

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US2005/010912

International filing date: 31 March 2005 (31.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/559,040
Filing date: 01 April 2004 (01.04.2004)

Date of receipt at the International Bureau: 12 August 2005 (12.08.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

1352393

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 02, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: 60/559,040

FILING DATE: April 01, 2004

RELATED PCT APPLICATION NUMBER: PCT/US05/10912



Certified by

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

15750 U.S. PTO
040104

PTO/SB/16 (08-03)
Approved for use through 7/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. EL 961008081 US

22858 U.S. PTO
60/559040

040104

INVENTOR(S)				
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)		
Steven	MAH	San Diego, California		
Andreas	BRAUN	San Diego, California		
Stefan M.	KAMMERER	San Diego, California		
Additional inventors are being named on the <u>1</u> separately numbered sheets attached hereto				
TITLE OF THE INVENTION (500 characters max)				
METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF				
Direct all correspondence to: CORRESPONDENCE ADDRESS				
<input checked="" type="checkbox"/> Customer Number: 25225				
OR				
<input type="checkbox"/> Firm or Individual Name				
Address				
City		State	Zip	
Country		Telephone	Fax	
ENCLOSED APPLICATION PARTS (check all that apply)				
<input checked="" type="checkbox"/> Specification Number of Pages		108	<input type="checkbox"/> CD(s), Number	
<input checked="" type="checkbox"/> Drawing(s) Number of Sheets		1	<input checked="" type="checkbox"/> Other	
<input checked="" type="checkbox"/> Application Data Sheet. See 37 CFR 1.76 (4 pages)		(specify): Return Receipt Postcard		
METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT				
<input checked="" type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27.				
<input type="checkbox"/> A check or money order is enclosed to cover the filing fees.				
<input checked="" type="checkbox"/> The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 03-1952				
<input type="checkbox"/> Payment by credit card. Form PTO-2038 is attached.				
FILING FEE AMOUNT (\$) 80.00				
The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.				
<input checked="" type="checkbox"/> No <input type="checkbox"/> Yes, the name of the U.S. Government agency and the Government contract number are:				

[Page 1 of 2]

Respectfully submitted,

Date April 1, 2004

SIGNATURE

TYPED OR

PRINTED NAME

Bruce D. Grant

REGISTRATION NO.

(if appropriate)

47,608

TELEPHONE

(858) 720-7962

Docket Number:

524593009100

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as Express Mail, Airbill No. EL 961008081 US, in an envelope addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date shown below.

Dated:

4/1/04

Signature:

Deborah Wykes

(Deborah Wykes)

SD-189469

PROVISIONAL APPLICATION COVER SHEET
Additional Page

PTO/SB/16 (08-03)

Approved for use through 07/31/06. OMB 0651-0032

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Docket Number 524593009100

INVENTOR(S)/APPLICANT(S)		
Given Name (first and middle [if any])	Family or Surname	Residence (City and either State or Foreign Country)
Matthew Roberts	NELSON	San Marcos, California
Rikard Henry	RENELAND	San Diego, California
Maria L.	LANGDOWN	San Diego, California

METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF

Field of the Invention

[0001] The invention relates to genetic methods for identifying risk of osteoarthritis and treatments that specifically target such diseases.

Background

[0002] Osteoarthritis (OA) is a chronic disease usually affecting weight-bearing synovial joints. There are approximately 20 million Americans affected by OA and it is the leading cause of disability in the United States. In addition to extensive human suffering, OA also accounts for nearly all knee replacements and more than half of all hip replacements in the United States. Despite its prevalence, OA is poorly understood and there are few treatments available besides anti-inflammatory drugs and joint replacement.

[0003] Most commonly affecting middle-aged and older people, OA can range from very mild to very severe. It affects hands and weight-bearing joints such as knees, hips, feet and the back. Knee OA can be as disabling as any cardiovascular disease except stroke.

[0004] OA is characterized by the breakdown of cartilage in joints. Cartilage in joints cushions the ends of bones, and cartilage breakdown causes bones to rub against each other, causing pain and loss of movement. Type II collagen is the main component of cartilage, comprising 15-25% of the wet weight, approximately half the dry weight, and representing 90-95% of the total collagen content in the tissue. It forms fibrils that endow cartilage with tensile strength (Mayne, R. Arthritis Rheum. 32:241-246 (1989)).

Summary

[0005] It has been discovered that certain polymorphic variations in human genomic DNA are associated with osteoarthritis. In particular, polymorphic variants in a locus containing a *TM7SF3* region in human genomic DNA have been associated with risk of osteoarthritis.

[0006] Thus, featured herein are methods for identifying a subject at risk of osteoarthritis and/or a risk of osteoarthritis in a subject, which comprise detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in or around the loci described herein in a human nucleic acid sample. In an embodiment, two or more polymorphic variations are detected in two or more regions of which one is the *TM7SF3* region. In certain embodiments, 3 or more, or 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 or more polymorphic variants are detected.

[0007] Also featured are nucleic acids that include one or more polymorphic variations associated with occurrence of osteoarthritis, as well as polypeptides encoded by these nucleic acids. In addition, provided are methods for identifying candidate therapeutic molecules for treating osteoarthritis, as well as methods for treating osteoarthritis in a subject by identifying a subject at risk of osteoarthritis and treating the subject with a suitable prophylactic, treatment or therapeutic molecule.

[0008] Also provided are compositions comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and/or a *TM7SF3* nucleic acid, with a RNAi, siRNA, antisense DNA or RNA, or ribozyme nucleic acid designed from a *TM7SF3* nucleotide sequence. In an embodiment, the RNAi, siRNA, antisense DNA or RNA, or ribozyme nucleic acid is designed from a *TM7SF3* nucleotide sequence that includes one or more polymorphic variations associated with osteoarthritis, and in some instances, specifically interacts with such a nucleotide sequence. Further, provided are arrays of nucleic acids bound to a solid surface, in which one or more nucleic acid molecules of the array have a *TM7SF3* nucleotide sequence, or a fragment or substantially identical nucleic acid thereof, or a complementary nucleic acid of the foregoing. Featured also are compositions comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and/or a *TM7SF3* polypeptide, with an antibody that specifically binds to the polypeptide. In an embodiment, the antibody specifically binds to an epitope in the polypeptide that includes a non-synonymous amino acid modification associated with osteoarthritis (e.g., results in an amino acid substitution in the encoded polypeptide associated with osteoarthritis).

Brief Description of the Drawings

[0009] Figure 1 shows proximal SNPs in a *TM7SF3* region in genomic DNA. The position of each SNP in the chromosome is shown on the x-axis and the y-axis provides the negative logarithm of the p-value comparing the estimated allele to that of the control group. Also shown in the figure are exons and introns of the gene region in the approximate chromosomal positions.

Detailed Description

[0010] It has been discovered that a polymorphic variant in a locus containing a *TM7SF3* region is associated with occurrence of osteoarthritis in subjects. Thus, detecting genetic determinants associated with an increased risk of osteoarthritis occurrence can lead to early identification of a predisposition to osteoarthritis and early prescription of preventative measures. Also, associating a *TM7SF3* polymorphic variant with osteoarthritis has provided new targets for screening molecules useful in treatments of osteoarthritis.

Osteoarthritis and Sample Selection

[0011] Osteoarthritis (OA), or degenerative joint disease, is one of the oldest and most common types of arthritis. It is characterized by the breakdown of the joint's cartilage. Cartilage is the part of the joint that cushions the ends of bones, and its breakdown causes bones to rub against each other, causing pain and loss of movement. Type II collagen is the main component of cartilage, comprising 15-25% of the wet weight, approximately half the dry weight, and representing 90-95% of the total collagen content in the tissue. It forms fibrils that endow cartilage with tensile strength (Mayne, R. Arthritis Rheum. 32:241-246 (1989)).

[0012] Most commonly affecting middle-aged and older people, OA can range from very mild to very severe. It affects hands and weight-bearing joints such as knees, hips, feet and the back. Knee OA can be as disabling as any cardiovascular disease except stroke.

[0013] Osteoarthritis affects an estimated 20.7 million Americans, mostly after age 45, with women more commonly affected than men. Physicians make a diagnosis of OA based on a physical exam and history of symptoms. X-rays are used to confirm diagnosis. Most people over 60 reflect the disease on X-ray, and about one-third have actual symptoms.

[0014] There are many factors that can cause OA. Obesity may lead to osteoarthritis of the knees. In addition, people with joint injuries due to sports, work-related activity or accidents may be at increased risk of developing OA.

[0015] Genetics has a role in the development of OA. Some people may be born with defective cartilage or with slight defects in the way that joints fit together. As a person ages, these defects may cause early cartilage breakdown in the joint or the inability to repair damaged or deteriorated cartilage in the joint.

[0016] Inclusion or exclusion of samples for an osteoarthritis pool may be based upon the following criteria: ethnicity (e.g., samples derived from an individual characterized as Caucasian); parental ethnicity (e.g., samples derived from an individual of British paternal and maternal descent); relevant phenotype information for the individual (e.g., case samples derived from individuals diagnosed with specific knee osteoarthritis (OA) and were recruited from an OA knee replacement clinic). Control samples may be selected based on relevant phenotype information for the individual (e.g., derived from individuals free of OA at several sites (knee, hand, hip etc)); and no family history of OA and/or rheumatoid arthritis. Additional phenotype information collected for both cases and controls may include age of the individual, gender, family history of OA, diagnosis with osteoarthritis (joint location of OA, date of primary diagnosis, age of individual as of primary diagnosis), knee history (current symptoms, any major knee injury, meniscectomy, knee replacement surgery, age of surgery), HRT history, osteoporosis diagnosis.

[0017] Based in part upon selection criteria set forth above, individuals having osteoarthritis can be selected for genetic studies. Also, individuals having no history of osteoarthritis often are selected for genetic studies, as described hereafter.

Polymorphic Variants Associated with Osteoarthritis

[0018] A genetic analysis provided herein linked osteoarthritis with polymorphic variant nucleic acid sequences in the human genome. As used herein, the term “polymorphic site” refers to a region in a nucleic acid at which two or more alternative nucleotide sequences are observed in a significant number of nucleic acid samples from a population of individuals. A polymorphic site may be a nucleotide sequence of two or more nucleotides, an inserted nucleotide or nucleotide sequence, a deleted nucleotide or nucleotide sequence, or a microsatellite, for example. A polymorphic site that is two or more nucleotides in length may be 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more, 20 or more, 30 or more, 50 or more, 75 or more, 100 or more, 500 or more, or about 1000 nucleotides in length, where all or some of the nucleotide sequences differ within the region. A polymorphic site is often one nucleotide in length, which is referred to herein as a “single nucleotide polymorphism” or a “SNP.”

[0019] Where there are two, three, or four alternative nucleotide sequences at a polymorphic site, each nucleotide sequence is referred to as a “polymorphic variant” or “nucleic acid variant.” Where two polymorphic variants exist, for example, the polymorphic variant represented in a minority of samples from a population is sometimes referred to as a “minor allele” and the polymorphic variant that is more prevalently represented is sometimes referred to as a “major allele.” Many organisms possess a copy of each chromosome (*e.g.*, humans), and those individuals who possess two major alleles or two minor alleles are often referred to as being “homozygous” with respect to the polymorphism, and those individuals who possess one major allele and one minor allele are normally referred to as being “heterozygous” with respect to the polymorphism. Individuals who are homozygous with respect to one allele are sometimes predisposed to a different phenotype as compared to individuals who are heterozygous or homozygous with respect to another allele.

[0020] In genetic analysis that associate polymorphic variants with osteoarthritis, samples from individuals having osteoarthritis and individuals not having osteoarthritis often are allelotyped and/or genotyped. The term “allelotype” as used herein refers to a process for determining the allele frequency for a polymorphic variant in pooled DNA samples from cases and controls. By pooling DNA from each group, an allele frequency for each SNP in each group is calculated. These allele frequencies are then compared to one another. The term “genotyped” as used herein refers to a process for determining a genotype of one or more individuals, where a “genotype” is a representation of one or more polymorphic variants in a population.

[0021] A genotype or polymorphic variant may be expressed in terms of a “haplotype,” which as used herein refers to two or more polymorphic variants occurring within genomic DNA in a group of individuals within a population. For example, two SNPs may exist within a gene where each SNP position includes a cytosine variation and an adenine variation. Certain individuals in a population may carry one allele (heterozygous) or two alleles (homozygous) having the gene with a cytosine at each SNP position. As the two cytosines corresponding to each SNP in the gene travel together on one or both alleles in these individuals, the individuals can be characterized as having a cytosine/cytosine haplotype with respect to the two SNPs in the gene.

[0022] As used herein, the term “phenotype” refers to a trait which can be compared between individuals, such as presence or absence of a condition, a visually observable difference in appearance between individuals, metabolic variations, physiological variations, variations in the function of biological molecules, and the like. An example of a phenotype is occurrence of osteoarthritis.

[0023] Researchers sometimes report a polymorphic variant in a database without determining whether the variant is represented in a significant fraction of a population. Because a subset of these reported polymorphic variants are not represented in a statistically significant portion of the population, some of them are sequencing errors and/or not biologically relevant. Thus, it is often not known whether a reported polymorphic variant is statistically significant or biologically relevant until the presence of the variant is detected in a population of individuals and the frequency of the variant is determined. Methods for detecting a polymorphic variant in a population are described herein, specifically in Example 2. A polymorphic variant is statistically significant and often biologically relevant if it is represented in 5% or more of a population, sometimes 10% or more, 15% or more, or 20% or more of a population, and often 25% or more, 30% or more, 35% or more, 40% or more, 45% or more, or 50% or more of a population.

[0024] A polymorphic variant may be detected on either or both strands of a double-stranded nucleic acid. Also, a polymorphic variant may be located within an intron or exon of a gene or within a portion of a regulatory region such as a promoter, a 5′ untranslated region (UTR), a 3′ UTR, and in DNA (*e.g.*, genomic DNA (gDNA) and complementary DNA (cDNA)), RNA (*e.g.*, mRNA, tRNA, and rRNA), or a polypeptide. Polymorphic variations may or may not result in detectable differences in gene expression, polypeptide structure, or polypeptide function.

[0025] It was determined that polymorphic variations associated with an increased risk of osteoarthritis existed in a *TM7SF3* region in SEQ ID NO: 1. In certain embodiments, a polymorphic variant at position rs1484086 in the human genome was associated with an increased risk of osteoarthritis, and in a specific embodiment, a thymine at position rs1484086 was associated with an increased risk of osteoarthritis.

[0026] Polymorphic variants in and around the *TM7SF3* locus were tested for association with osteoarthritis. These include polymorphic variants at positions in SEQ ID NO: 1 selected from the group

consisting of 230, 231, 5330, 6334, 11372, 11456, 11501, 13393, 16666, 17596, 19710, 19800, 20297, 20967, 32514, 33159, 37600, 41259, 41329, 50060, 53292, 53393, 56417, 56435, 58847, 59595, 59661, 60355, 60407, 62357, 68230, 68516, 69055, 72603, 73928, 85897 and 91554. Polymorphic variants at the following positions in SEQ ID NO: 1 in particular were associated with an increased risk of osteoarthritis: 56435, 59595, 53292, 33159 and 41329, with specific embodiments directed to positions 56435 and/or 59595. In particular, the following polymorphic variants in SEQ ID NO: 1 were associated with risk of osteoarthritis: a thymine thymine repeat at position 56435, a thymine at position 59595, a cytosine at position 53292, a guanine at position 33159 and a thymine at position 41329.

[0027] Based in part upon analyses summarized in Figure 1, a region with significant association has been identified in a locus associated with osteoarthritis. Any polymorphic variants associated with osteoarthritis in a region of significant association can be utilized for embodiments described herein. For example, polymorphic variants in a region approximately 14,000 nucleotides in length spanning chromosome positions 27052000 to 27066000 in a *TM7SF3* locus have significant association (chromosome positions are within NCBI's Genome build 34).

Additional Polymorphic Variants Associated with Osteoarthritis

[0028] Also provided is a method for identifying polymorphic variants proximal to an incident, founder polymorphic variant associated with osteoarthritis. Thus, featured herein are methods for identifying a polymorphic variation associated with osteoarthritis that is proximal to an incident polymorphic variation associated with osteoarthritis, which comprises identifying a polymorphic variant proximal to the incident polymorphic variant associated with osteoarthritis, where the incident polymorphic variant is in a *TM7SF3* nucleotide sequence. The nucleotide sequence often comprises a polynucleotide sequence selected from the group consisting of (a) a polynucleotide sequence of SEQ ID NO: 1-2; (b) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence encoded by a polynucleotide sequence of SEQ ID NO: 1-2; and (c) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2 or a polynucleotide sequence 90% or more identical to the polynucleotide sequence of SEQ ID NO: 1-2. The presence or absence of an association of the proximal polymorphic variant with osteoarthritis then is determined using a known association method, such as a method described in the Examples hereafter. In an embodiment, the incident polymorphic variant is a polymorphic variant associated with osteoarthritis described herein. In another embodiment, the proximal polymorphic variant identified sometimes is a publicly disclosed polymorphic variant, which for example, sometimes is published in a publicly available database. In other embodiments, the polymorphic variant identified is not publicly disclosed and is discovered using a known method, including, but not limited to, sequencing a region surrounding the incident polymorphic

variant in a group of nucleic samples. Thus, multiple polymorphic variants proximal to an incident polymorphic variant are associated with osteoarthritis using this method.

[0029] The proximal polymorphic variant often is identified in a region surrounding the incident polymorphic variant. In certain embodiments, this surrounding region is about 50 kb flanking the first polymorphic variant (*e.g.* about 50 kb 5' of the first polymorphic variant and about 50 kb 3' of the first polymorphic variant), and the region sometimes is composed of shorter flanking sequences, such as flanking sequences of about 40 kb, about 30 kb, about 25 kb, about 20 kb, about 15 kb, about 10 kb, about 7 kb, about 5 kb, or about 2 kb 5' and 3' of the incident polymorphic variant. In other embodiments, the region is composed of longer flanking sequences, such as flanking sequences of about 55 kb, about 60 kb, about 65 kb, about 70 kb, about 75 kb, about 80 kb, about 85 kb, about 90 kb, about 95 kb, or about 100 kb 5' and 3' of the incident polymorphic variant.

[0030] In certain embodiments, polymorphic variants associated with osteoarthritis are identified iteratively. For example, a first proximal polymorphic variant is associated with osteoarthritis using the methods described above and then another polymorphic variant proximal to the first proximal polymorphic variant is identified (*e.g.*, publicly disclosed or discovered) and the presence or absence of an association of one or more other polymorphic variants proximal to the first proximal polymorphic variant with osteoarthritis is determined.

[0031] The methods described herein are useful for identifying or discovering additional polymorphic variants that may be used to further characterize a gene, region or loci associated with a condition, a disease (*e.g.*, osteoarthritis), or a disorder. For example, allelotyping or genotyping data from the additional polymorphic variants may be used to identify a functional mutation or a region of linkage disequilibrium. In certain embodiments, polymorphic variants identified or discovered within a region comprising the first polymorphic variant associated with osteoarthritis are genotyped using the genetic methods and sample selection techniques described herein, and it can be determined whether those polymorphic variants are in linkage disequilibrium with the first polymorphic variant. The size of the region in linkage disequilibrium with the first polymorphic variant also can be assessed using these genotyping methods. Thus, provided herein are methods for determining whether a polymorphic variant is in linkage disequilibrium with a first polymorphic variant associated with osteoarthritis, and such information can be used in prognosis/diagnosis methods described herein.

Isolated Nucleic Acids

[0032] Featured herein are isolated *TM7SF3* nucleic acid variants depicted in SEQ ID NO: 1-2, and substantially identical nucleic acids thereof. A nucleic acid variant may be represented on one or both strands in a double-stranded nucleic acid or on one chromosomal complement (heterozygous) or both chromosomal complements (homozygous).

[0033] As used herein, the term “nucleic acid” includes DNA molecules (*e.g.*, a complementary DNA (cDNA) and genomic DNA (gDNA)) and RNA molecules (*e.g.*, mRNA, rRNA, siRNA and tRNA) and analogs of DNA or RNA, for example, by use of nucleotide analogs. The nucleic acid molecule can be single-stranded and it is often double-stranded. The term “isolated or purified nucleic acid” refers to nucleic acids that are separated from other nucleic acids present in the natural source of the nucleic acid. For example, with regard to genomic DNA, the term “isolated” includes nucleic acids which are separated from the chromosome with which the genomic DNA is naturally associated. An “isolated” nucleic acid is often free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5’ and/or 3’ ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5’ and/or 3’ nucleotide sequences which flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an “isolated” nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. As used herein, the term “gene” refers to a nucleotide sequence that encodes a polypeptide.

[0034] Also included herein are nucleic acid fragments. These fragments often have a nucleotide sequence identical to a nucleotide sequence of SEQ ID NO: 1-2, a nucleotide sequence substantially identical to a nucleotide sequence of SEQ ID NO: 1-2, or a nucleotide sequence that is complementary to the foregoing. The nucleic acid fragment may be identical, substantially identical or homologous to a nucleotide sequence in an exon or an intron in a nucleotide sequence of SEQ ID NO: 1-2, and may encode a domain or part of a domain of a polypeptide. Sometimes, the fragment will comprises one or more of the polymorphic variations described herein as being associated with osteoarthritis. The nucleic acid fragment is often 50, 100, or 200 or fewer base pairs in length, and is sometimes about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 2000, 3000, 4000, 5000, 10000, 15000, or 20000 base pairs in length. A nucleic acid fragment that is complementary to a nucleotide sequence identical or substantially identical to a nucleotide sequence in SEQ ID NO: 1-2 and hybridizes to such a nucleotide sequence under stringent conditions is often referred to as a “probe.” Nucleic acid fragments often include one or more polymorphic sites, or sometimes have an end that is adjacent to a polymorphic site as described hereafter.

[0035] An example of a nucleic acid fragment is an oligonucleotide. As used herein, the term “oligonucleotide” refers to a nucleic acid comprising about 8 to about 50 covalently linked nucleotides, often comprising from about 8 to about 35 nucleotides, and more often from about 10 to about 25 nucleotides. The backbone and nucleotides within an oligonucleotide may be the same as those of naturally occurring nucleic acids, or analogs or derivatives of naturally occurring nucleic acids, provided

that oligonucleotides having such analogs or derivatives retain the ability to hybridize specifically to a nucleic acid comprising a targeted polymorphism. Oligonucleotides described herein may be used as hybridization probes or as components of prognostic or diagnostic assays, for example, as described herein.

[0036] Oligonucleotides are typically synthesized using standard methods and equipment, such as the ABI™3900 High Throughput DNA Synthesizer and the EXPEDITE™ 8909 Nucleic Acid Synthesizer, both of which are available from Applied Biosystems (Foster City, CA). Analogs and derivatives are exemplified in U.S. Pat. Nos. 4,469,863; 5,536,821; 5,541,306; 5,637,683; 5,637,684; 5,700,922; 5,717,083; 5,719,262; 5,739,308; 5,773,601; 5,886,165; 5,929,226; 5,977,296; 6,140,482; WO 00/56746; WO 01/14398, and related publications. Methods for synthesizing oligonucleotides comprising such analogs or derivatives are disclosed, for example, in the patent publications cited above and in U.S. Pat. Nos. 5,614,622; 5,739,314; 5,955,599; 5,962,674; 6,117,992; in WO 00/75372; and in related publications.

[0037] Oligonucleotides may also be linked to a second moiety. The second moiety may be an additional nucleotide sequence such as a tail sequence (*e.g.*, a polyadenosine tail), an adapter sequence (*e.g.*, phage M13 universal tail sequence), and others. Alternatively, the second moiety may be a non-nucleotide moiety such as a moiety which facilitates linkage to a solid support or a label to facilitate detection of the oligonucleotide. Such labels include, without limitation, a radioactive label, a fluorescent label, a chemiluminescent label, a paramagnetic label, and the like. The second moiety may be attached to any position of the oligonucleotide, provided the oligonucleotide can hybridize to the nucleic acid comprising the polymorphism.

Uses for Nucleic Acid Sequence

[0038] Nucleic acid coding sequences may be used for diagnostic purposes for detection and control of polypeptide expression. Also, included herein are oligonucleotide sequences such as antisense RNA, small-interfering RNA (siRNA) and DNA molecules and ribozymes that function to inhibit translation of a polypeptide. Antisense techniques and RNA interference techniques are known in the art and are described herein.

[0039] Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, hammerhead motif ribozyme molecules may be engineered that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences corresponding to or complementary to *TM7SF3* nucleotide sequences. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA,

GUU and GUC. Once identified, short RNA sequences of between fifteen (15) and twenty (20) ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features such as secondary structure that may render the oligonucleotide sequence unsuitable. The suitability of candidate targets may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

[0040] Antisense RNA and DNA molecules, siRNA and ribozymes may be prepared by any method known in the art for the synthesis of RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides well known in the art such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

[0041] DNA encoding a polypeptide also may have a number of uses for the diagnosis of diseases, including osteoarthritis, resulting from aberrant expression of a target gene described herein. For example, the nucleic acid sequence may be used in hybridization assays of biopsies or autopsies to diagnose abnormalities of expression or function (*e.g.*, Southern or Northern blot analysis, *in situ* hybridization assays).

[0042] In addition, the expression of a polypeptide during embryonic development may also be determined using nucleic acid encoding the polypeptide. As addressed, *infra*, production of functionally impaired polypeptide is the cause of various disease states, such as osteoarthritis. *In situ* hybridizations using polypeptide as a probe may be employed to predict problems related to osteoarthritis. Further, as indicated, *infra*, administration of human active polypeptide, recombinantly produced as described herein, may be used to treat disease states related to functionally impaired polypeptide. Alternatively, gene therapy approaches may be employed to remedy deficiencies of functional polypeptide or to replace or compete with dysfunctional polypeptide.

Expression Vectors, Host Cells, and Genetically Engineered Cells

[0043] Provided herein are nucleic acid vectors, often expression vectors, which contain a *TM7SF3* nucleotide sequence, or a substantially identical sequence thereof. As used herein, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid, or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors may include replication defective retroviruses, adenoviruses and adeno-associated viruses for example.

[0044] A vector can include a *TM7SF3* nucleotide sequence in a form suitable for expression of an encoded target polypeptide or target nucleic acid in a host cell. A “target polypeptide” is a polypeptide encoded by a *TM7SF3* nucleotide sequence, or a substantially identical nucleotide sequence thereof. The recombinant expression vector typically includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term “regulatory sequence” includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, and the like. Expression vectors can be introduced into host cells to produce target polypeptides, including fusion polypeptides.

[0045] Recombinant expression vectors can be designed for expression of target polypeptides in prokaryotic or eukaryotic cells. For example, target polypeptides can be expressed in *E. coli*, insect cells (*e.g.*, using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology 185*, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[0046] Expression of polypeptides in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polypeptides. Fusion vectors add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant polypeptide; 2) to increase the solubility of the recombinant polypeptide; and 3) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide to enable separation of the recombinant polypeptide from the fusion moiety subsequent to purification of the fusion polypeptide. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith & Johnson, *Gene* 67: 31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding polypeptide, or polypeptide A, respectively, to the target recombinant polypeptide.

[0047] Purified fusion polypeptides can be used in screening assays and to generate antibodies specific for target polypeptides. In a therapeutic embodiment, fusion polypeptide expressed in a retroviral expression vector is used to infect bone marrow cells that are subsequently transplanted into

irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (*e.g.*, six (6) weeks).

[0048] Expressing the polypeptide in host bacteria with an impaired capacity to proteolytically cleave the recombinant polypeptide is often used to maximize recombinant polypeptide expression (Gottesman, S., *Gene Expression Technology: Methods in Enzymology*, Academic Press, San Diego, California 185: 119-128 (1990)). Another strategy is to alter the nucleotide sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, *Nucleic Acids Res.* 20: 2111-2118 (1992)). Such alteration of nucleotide sequences can be carried out by standard DNA synthesis techniques.

[0049] When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. Recombinant mammalian expression vectors are often capable of directing expression of the nucleic acid in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include an albumin promoter (liver-specific; Pinkert *et al.*, *Genes Dev.* 1: 268-277 (1987)), lymphoid-specific promoters (Calame & Eaton, *Adv. Immunol.* 43: 235-275 (1988)), promoters of T cell receptors (Winoto & Baltimore, *EMBO J.* 8: 729-733 (1989)) promoters of immunoglobulins (Banerji *et al.*, *Cell* 33: 729-740 (1983); Queen & Baltimore, *Cell* 33: 741-748 (1983)), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne & Ruddle, *Proc. Natl. Acad. Sci. USA* 86: 5473-5477 (1989)), pancreas-specific promoters (Edlund *et al.*, *Science* 230: 912-916 (1985)), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are sometimes utilized, for example, the murine hox promoters (Kessel & Gruss, *Science* 249: 374-379 (1990)) and the α -fetopolypeptide promoter (Campes & Tilghman, *Genes Dev.* 3: 537-546 (1989)).

[0050] A *TM7SF3* nucleic acid also may be cloned into an expression vector in an antisense orientation. Regulatory sequences (*e.g.*, viral promoters and/or enhancers) operatively linked to a *TM7SF3* nucleic acid cloned in the antisense orientation can be chosen for directing constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. Antisense expression vectors can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes *see, e.g.*, Weintraub *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) (1986).

[0051] Also provided herein are host cells that include a *TM7SF3* nucleotide sequence within a recombinant expression vector or a fragment of such a nucleotide sequence which facilitate homologous recombination into a specific site of the host cell genome. The terms "host cell" and "recombinant host

cell” are used interchangeably herein. Such terms refer not only to the particular subject cell but rather also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a target polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[0052] Vectors can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms “transformation” and “transfection” are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, transduction/infection, DEAE-dextran-mediated transfection, lipofection, or electroporation.

[0053] A host cell provided herein can be used to produce (*i.e.*, express) a target polypeptide or a substantially identical polypeptide thereof. Accordingly, further provided are methods for producing a target polypeptide using host cells described herein. In one embodiment, the method includes culturing host cells into which a recombinant expression vector encoding a target polypeptide has been introduced in a suitable medium such that a target polypeptide is produced. In another embodiment, the method further includes isolating a target polypeptide from the medium or the host cell.

[0054] Also provided are cells or purified preparations of cells which include a *TM7SF3* transgene, or which otherwise misexpress target polypeptide. Cell preparations can consist of human or non-human cells, *e.g.*, rodent cells, *e.g.*, mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a *TM7SF3* transgene (*e.g.*, a heterologous form of a *TM7SF3* gene, such as a human gene expressed in non-human cells). The transgene can be misexpressed, *e.g.*, overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpress an endogenous target polypeptide (*e.g.*, expression of a gene is disrupted, also known as a knockout). Such cells can serve as a model for studying disorders which are related to mutated or mis-expressed alleles or for use in drug screening. Also provided are human cells (*e.g.*, a hematopoietic stem cells) transfected with a *TM7SF3* nucleic acid.

[0055] Also provided are cells or a purified preparation thereof (*e.g.*, human cells) in which an endogenous *TM7SF3* nucleic acid is under the control of a regulatory sequence that does not normally control the expression of the endogenous gene. The expression characteristics of an endogenous gene within a cell (*e.g.*, a cell line or microorganism) can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the corresponding endogenous gene. For example, an endogenous corresponding gene (*e.g.*, a gene which is “transcriptionally silent,” not normally expressed, or expressed only at very low levels)

may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, *e.g.*, Chappel, US 5,272,071; WO 91/06667, published on May 16, 1991.

Transgenic Animals

[0056] Non-human transgenic animals that express a heterologous target polypeptide (*e.g.*, expressed from a *TM7SF3* nucleic acid or substantially identical sequence thereof) can be generated. Such animals are useful for studying the function and/or activity of a target polypeptide and for identifying and/or evaluating modulators of the activity of *TM7SF3* nucleic acids and encoded polypeptides. As used herein, a “transgenic animal” is a non-human animal such as a mammal (*e.g.*, a non-human primate such as chimpanzee, baboon, or macaque; an ungulate such as an equine, bovine, or caprine; or a rodent such as a rat, a mouse, or an Israeli sand rat), a bird (*e.g.*, a chicken or a turkey), an amphibian (*e.g.*, a frog, salamander, or newt), or an insect (*e.g.*, *Drosophila melanogaster*), in which one or more of the cells of the animal includes a transgene. A transgene is exogenous DNA or a rearrangement (*e.g.*, a deletion of endogenous chromosomal DNA) that is often integrated into or occurs in the genome of cells in a transgenic animal. A transgene can direct expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, and other transgenes can reduce expression (*e.g.*, a knockout). Thus, a transgenic animal can be one in which an endogenous nucleic acid homologous to a *TM7SF3* nucleic acid has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal (*e.g.*, an embryonic cell of the animal) prior to development of the animal.

[0057] Intronic sequences and polyadenylation signals can also be included in the transgene to increase expression efficiency of the transgene. One or more tissue-specific regulatory sequences can be operably linked to a *TM7SF3* nucleotide sequence to direct expression of an encoded polypeptide to particular cells. A transgenic founder animal can be identified based upon the presence of a *TM7SF3* nucleotide sequence in its genome and/or expression of encoded mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a *TM7SF3* nucleotide sequence can further be bred to other transgenic animals carrying other transgenes.

[0058] Target polypeptides can be expressed in transgenic animals or plants by introducing, for example, a *TM7SF3* nucleic acid into the genome of an animal that encodes the target polypeptide. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, *e.g.*, a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Also included is a population of cells from a transgenic animal.

Target Polypeptides

[0059] Also featured herein are isolated target polypeptides, which are encoded by a *TM7SF3* nucleotide sequence (e.g., SEQ ID NO: 1-2), or a substantially identical nucleotide sequence thereof. Examples of *TM7SF3* polypeptides are set forth in SEQ ID NO: 3. The term “polypeptide” as used herein includes proteins and peptides. An “isolated” or “purified” polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language “substantially free” means preparation of a target polypeptide having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-target polypeptide (also referred to herein as a “contaminating protein”), or of chemical precursors or non-target chemicals. When the target polypeptide or a biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, specifically, where culture medium represents less than about 20%, sometimes less than about 10%, and often less than about 5% of the volume of the polypeptide preparation. Isolated or purified target polypeptide preparations are sometimes 0.01 milligrams or more or 0.1 milligrams or more, and often 1.0 milligrams or more and 10 milligrams or more in dry weight.

[0060] Further included herein are target polypeptide fragments. The polypeptide fragment may be a domain or part of a domain of a target polypeptide. The polypeptide fragment may have increased, decreased or unexpected biological activity. The polypeptide fragment is often 50 or fewer, 100 or fewer, or 200 or fewer amino acids in length, and is sometimes 300, 400, 500, 600, 700, or 900 or fewer amino acids in length.

[0061] Substantially identical target polypeptides may depart from the amino acid sequences of target polypeptides in different manners. For example, conservative amino acid modifications may be introduced at one or more positions in the amino acid sequences of target polypeptides. A “conservative amino acid substitution” is one in which the amino acid is replaced by another amino acid having a similar structure and/or chemical function. Families of amino acid residues having similar structures and functions are well known. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Also, essential and non-essential amino acids may be replaced. A “non-essential” amino acid is one that can be altered without abolishing or substantially altering the biological function of a target polypeptide, whereas altering an “essential” amino acid abolishes or substantially alters the biological function of a target polypeptide. Amino acids that are conserved among target polypeptides

are typically essential amino acids. In certain embodiments, the polypeptide includes one or more non-synonymous polymorphic variants associated with osteoarthritis.

[0062] Also, target polypeptides may exist as chimeric or fusion polypeptides. As used herein, a target “chimeric polypeptide” or target “fusion polypeptide” includes a target polypeptide linked to a non-target polypeptide. A “non-target polypeptide” refers to a polypeptide having an amino acid sequence corresponding to a polypeptide which is not substantially identical to the target polypeptide, which includes, for example, a polypeptide that is different from the target polypeptide and derived from the same or a different organism. The target polypeptide in the fusion polypeptide can correspond to an entire or nearly entire target polypeptide or a fragment thereof. The non-target polypeptide can be fused to the N-terminus or C-terminus of the target polypeptide.

[0063] Fusion polypeptides can include a moiety having high affinity for a ligand. For example, the fusion polypeptide can be a GST-target fusion polypeptide in which the target sequences are fused to the C-terminus of the GST sequences, or a polyhistidine-target fusion polypeptide in which the target polypeptide is fused at the N- or C-terminus to a string of histidine residues. Such fusion polypeptides can facilitate purification of recombinant target polypeptide. Expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide), and a nucleotide sequence in SEQ ID NO: 1-2, or a substantially identical nucleotide sequence thereof, can be cloned into an expression vector such that the fusion moiety is linked in-frame to the target polypeptide. Further, the fusion polypeptide can be a target polypeptide containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression, secretion, cellular internalization, and cellular localization of a target polypeptide can be increased through use of a heterologous signal sequence. Fusion polypeptides can also include all or a part of a serum polypeptide (*e.g.*, an IgG constant region or human serum albumin).

[0064] Target polypeptides can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. Administration of these target polypeptides can be used to affect the bioavailability of a substrate of the target polypeptide and may effectively increase target polypeptide biological activity in a cell. Target fusion polypeptides may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a target polypeptide; (ii) mis-regulation of the gene encoding the target polypeptide; and (iii) aberrant post-translational modification of a target polypeptide. Also, target polypeptides can be used as immunogens to produce anti-target antibodies in a subject, to purify target polypeptide ligands or binding partners, and in screening assays to identify molecules which inhibit or enhance the interaction of a target polypeptide with a substrate.

[0065] In addition, polypeptides can be chemically synthesized using techniques known in the art (See, *e.g.*, Creighton, 1983 *Proteins*. New York, N.Y.: W. H. Freeman and Company; and Hunkapiller et al., (1984) *Nature* July 12 -18;310(5973):105-11). For example, a relative short fragment can be

synthesized by use of a peptide synthesizer. Furthermore, if desired, non-classical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the fragment sequence. Non-classical amino acids include, but are not limited to, to the D-isomers of the common amino acids, 2,4-diaminobutyric acid, α -amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, g-Abu, e-Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, homocitrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, b-alanine, fluoroamino acids, designer amino acids such as b-methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

[0066] Polypeptides and polypeptide fragments sometimes are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including but not limited, to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH₄; acetylation, formylation, oxidation, reduction; metabolic synthesis in the presence of tunicamycin; and the like. Additional post-translational modifications include, for example, N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends), attachment of chemical moieties to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, and addition or deletion of an N-terminal methionine residue as a result of prokaryotic host cell expression. The polypeptide fragments may also be modified with a detectable label, such as an enzymatic, fluorescent, isotopic or affinity label to allow for detection and isolation of the polypeptide.

[0067] Also provided are chemically modified derivatives of polypeptides that can provide additional advantages such as increased solubility, stability and circulating time of the polypeptide, or decreased immunogenicity (*see e.g.*, U.S. Pat. No: 4,179,337. The chemical moieties for derivitization may be selected from water soluble polymers such as polyethylene glycol, ethylene glycol/propylene glycol copolymers, carboxymethylcellulose, dextran, polyvinyl alcohol and the like. The polypeptides may be modified at random positions within the molecule, or at predetermined positions within the molecule and may include one, two, three or more attached chemical moieties.

[0068] The polymer may be of any molecular weight, and may be branched or unbranched. For polyethylene glycol, the preferred molecular weight is between about 1 kDa and about 100 kDa (the term "about" indicating that in preparations of polyethylene glycol, some molecules will weigh more, some less, than the stated molecular weight) for ease in handling and manufacturing. Other sizes may be used, depending on the desired therapeutic profile (*e.g.*, the duration of sustained release desired, the effects, if any on biological activity, the ease in handling, the degree or lack of antigenicity and other known effects of the polyethylene glycol to a therapeutic protein or analog).

[0069] The polymers should be attached to the polypeptide with consideration of effects on functional or antigenic domains of the polypeptide. There are a number of attachment methods available to those skilled in the art (*e.g.*, EP 0 401 384 (coupling PEG to G-CSF) and Malik et al. (1992) *Exp Hematol.* September;20(8):1028-35 (pegylation of GM-CSF using tresyl chloride)). For example, polyethylene glycol may be covalently bound through amino acid residues via a reactive group, such as a free amino or carboxyl group. Reactive groups are those to which an activated polyethylene glycol molecule may be bound. The amino acid residues having a free amino group may include lysine residues and the N-terminal amino acid residues; those having a free carboxyl group may include aspartic acid residues, glutamic acid residues and the C-terminal amino acid residue. Sulfhydryl groups may also be used as a reactive group for attaching the polyethylene glycol molecules. For therapeutic purposes, the attachment sometimes is at an amino group, such as attachment at the N-terminus or lysine group.

[0070] Proteins can be chemically modified at the N-terminus. Using polyethylene glycol as an illustration of such a composition, one may select from a variety of polyethylene glycol molecules (by molecular weight, branching, and the like), the proportion of polyethylene glycol molecules to protein (polypeptide) molecules in the reaction mix, the type of pegylation reaction to be performed, and the method of obtaining the selected N-terminally pegylated protein. The method of obtaining the N-terminally pegylated preparation (*i.e.*, separating this moiety from other monopegylated moieties if necessary) may be by purification of the N-terminally pegylated material from a population of pegylated protein molecules. Selective proteins chemically modified at the N-terminus may be accomplished by reductive alkylation, which exploits differential reactivity of different types of primary amino groups (lysine versus the N-terminal) available for derivatization in a particular protein. Under the appropriate reaction conditions, substantially selective derivatization of the protein at the N-terminus with a carbonyl group containing polymer is achieved.

Substantially Identical Nucleic Acids and Polypeptides

[0071] Nucleotide sequences and polypeptide sequences that are substantially identical to a *TM7SF3* nucleotide sequence and the target polypeptide sequences encoded by those nucleotide sequences, respectively, are included herein. The term “substantially identical” as used herein refers to two or more nucleic acids or polypeptides sharing one or more identical nucleotide sequences or polypeptide sequences, respectively. Included are nucleotide sequences or polypeptide sequences that are 55% or more, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, 95% or more (each often within a 1%, 2%, 3% or 4% variability) identical to a *TM7SF3* nucleotide sequence or the encoded target polypeptide amino acid sequences. One test for determining whether two nucleic acids are substantially identical is to determine the percent of identical nucleotide sequences or polypeptide sequences shared between the nucleic acids or polypeptides.

[0072] Calculations of sequence identity are often performed as follows. Sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is sometimes 30% or more, 40% or more, 50% or more, often 60% or more, and more often 70% or more, 80% or more, 90% or more, or 100% of the length of the reference sequence. The nucleotides or amino acids at corresponding nucleotide or polypeptide positions, respectively, are then compared among the two sequences. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, the nucleotides or amino acids are deemed to be identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, introduced for optimal alignment of the two sequences.

[0073] Comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. Percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers & Miller, *CABIOS* 4: 11-17 (1989), which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. Also, percent identity between two amino acid sequences can be determined using the Needleman & Wunsch, *J. Mol. Biol.* 48: 444-453 (1970) algorithm which has been incorporated into the GAP program in the GCG software package (available at the http address www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. Percent identity between two nucleotide sequences can be determined using the GAP program in the GCG software package (available at http address www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A set of parameters often used is a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[0074] Another manner for determining if two nucleic acids are substantially identical is to assess whether a polynucleotide homologous to one nucleic acid will hybridize to the other nucleic acid under stringent conditions. As use herein, the term "stringent conditions" refers to conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. An example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C,

followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Often, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. More often, stringency conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C.

[0075] An example of a substantially identical nucleotide sequence to a nucleotide sequence in SEQ ID NO: 1-2 is one that has a different nucleotide sequence but still encodes the same polypeptide sequence encoded by the nucleotide sequence in SEQ ID NO: 1-2. Another example is a nucleotide sequence that encodes a polypeptide having a polypeptide sequence that is more than 70% or more identical to, sometimes more than 75% or more, 80% or more, or 85% or more identical to, and often more than 90% or more and 95% or more identical to a polypeptide sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2.

[0076] Nucleotide sequences in SEQ ID NO: 1-2 and amino acid sequences of encoded polypeptides can be used as “query sequences” to perform a search against public databases to identify other family members or related sequences, for example. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.*, *J. Mol. Biol.* 215: 403-10 (1990). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleotide sequences in SEQ ID NO: 1-2. BLAST polypeptide searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to polypeptides encoded by the nucleotide sequences of SEQ ID NO: 1-2. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, *Nucleic Acids Res.* 25(17): 3389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used (*see* the http address www.ncbi.nlm.nih.gov).

[0077] A nucleic acid that is substantially identical to a nucleotide sequence in SEQ ID NO: 1-2 may include polymorphic sites at positions equivalent to those described herein when the sequences are aligned. For example, using the alignment procedures described herein, SNPs in a sequence substantially identical to a sequence in SEQ ID NO: 1-2 can be identified at nucleotide positions that match (*i.e.*, align) with nucleotides at SNP positions in each nucleotide sequence in SEQ ID NO: 1-2. Also, where a polymorphic variation results in an insertion or deletion, insertion or deletion of a nucleotide sequence from a reference sequence can change the relative positions of other polymorphic sites in the nucleotide sequence.

[0078] Substantially identical nucleotide and polypeptide sequences include those that are naturally occurring, such as allelic variants (same locus), splice variants, homologs (different locus), and orthologs (different organism) or can be non-naturally occurring. Non-naturally occurring variants can be generated by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product). Orthologs, homologs, allelic variants, and splice variants can be identified using methods known in the art. These variants normally comprise a nucleotide sequence encoding a polypeptide that is 50% or more, about 55% or more, often about 70-75% or more or about 80-85% or more, and sometimes about 90-95% or more identical to the amino acid sequences of target polypeptides or a fragment thereof. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions to a nucleotide sequence in SEQ ID NO: 1-2 or a fragment of this sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of a nucleotide sequence in SEQ ID NO: 1-2 can further be identified by mapping the sequence to the same chromosome or locus as the nucleotide sequence in SEQ ID NO: 1-2.

[0079] Also, substantially identical nucleotide sequences may include codons that are altered with respect to the naturally occurring sequence for enhancing expression of a target polypeptide in a particular expression system. For example, the nucleic acid can be one in which one or more codons are altered, and often 10% or more or 20% or more of the codons are altered for optimized expression in bacteria (*e.g.*, *E. coli.*), yeast (*e.g.*, *S. cerevisiae*), human (*e.g.*, 293 cells), insect, or rodent (*e.g.*, hamster) cells.

Methods for Identifying Risk of Osteoarthritis

[0080] Methods for prognosing and diagnosing osteoarthritis are included herein. These methods include detecting the presence or absence of one or more polymorphic variations in a nucleotide sequence associated with osteoarthritis, such as variants in or around the loci set forth herein, or a substantially identical sequence thereof, in a sample from a subject, where the presence of a polymorphic variant described herein is indicative of a risk of osteoarthritis. Determining a risk of osteoarthritis sometimes refers to determining whether an individual is at an increased risk of osteoarthritis (*e.g.*, intermediate risk or higher risk).

[0081] Thus, featured herein is a method for identifying a subject who is at risk of osteoarthritis, which comprises detecting an aberration associated with osteoarthritis in a nucleic acid sample from the subject. An embodiment is a method for detecting a risk of osteoarthritis in a subject, which comprises detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a

polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject, where the nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) a nucleotide sequence of SEQ ID NO: 1-2; (b) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-2; and (d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising the polymorphic site; whereby the presence of the polymorphic variation is indicative of a predisposition to osteoarthritis in the subject. In certain embodiments, polymorphic variants at the positions described herein are detected for determining a risk of osteoarthritis, and polymorphic variants at positions in linkage disequilibrium with these positions are detected for determining a risk of osteoarthritis. As used herein, "SEQ ID NO: 1-2" refers to individual sequences in SEQ ID NO: 1 or 2, each sequence being separately applicable to embodiments described herein.

[0082] Risk of osteoarthritis sometimes is expressed as a probability, such as an odds ratio, percentage, or risk factor. Risk often is based upon the presence or absence of one or more polymorphic variants described herein, and also may be based in part upon phenotypic traits of the individual being tested. Methods for calculating risk based upon patient data are well known (*see, e.g.,* Agresti, *Categorical Data Analysis*, 2nd Ed. 2002. Wiley). Allelotyping and genotyping analyses may be carried out in populations other than those exemplified herein to enhance the predictive power of the prognostic method. These further analyses are executed in view of the exemplified procedures described herein, and may be based upon the same polymorphic variations or additional polymorphic variations.

[0083] In certain embodiments, determining the presence of a combination of two or more polymorphic variants associated with osteoarthritis in one or more genetic loci (e.g., one or more genes) of the sample is determined to identify, quantify and/or estimate, risk of osteoarthritis. The risk often is the probability of having or developing osteoarthritis. The risk sometimes is expressed as a relative risk with respect to a population average risk of osteoarthritis, and sometimes is expressed as a relative risk with respect to the lowest risk group. Such relative risk assessments often are based upon penetrance values determined by statistical methods, and are particularly useful to clinicians and insurance companies for assessing risk of osteoarthritis (e.g., a clinician can target appropriate detection, prevention and therapeutic regimens to a patient after determining the patient's risk of osteoarthritis, and an insurance company can fine tune actuarial tables based upon population genotype assessments of osteoarthritis risk). Risk of osteoarthritis sometimes is expressed as an odds ratio, which is the odds of a particular person having a genotype has or will develop osteoarthritis with respect to another genotype group (e.g., the most disease protective genotype or population average). In related embodiments, the determination is utilized to identify a subject at risk of osteoarthritis. In an embodiment, two or more

polymorphic variations are detected in two or more regions in human genomic DNA associated with increased risk of osteoarthritis, such as a locus containing a *TM7SF3*, for example. In certain embodiments, 3 or more, or 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 or more polymorphic variants are detected in the sample. In specific embodiments, polymorphic variants are detected in a *TM7SF3* region, for example. In certain embodiments, polymorphic variants are detected at other genetic loci (e.g., the polymorphic variants can be detected in *TM7SF3* in addition to other loci or only in other loci), where the other loci include but are not limited to those described in concurrently-filed patent applications having attorney docket number 524593008700, 524593008800, 524593008900, 524593009000 or 524593009200, which is incorporated herein by reference in its entirety.

[0084] Results from prognostic tests may be combined with other test results to diagnose osteoarthritis. For example, prognostic results may be gathered, a patient sample may be ordered based on a determined predisposition to osteoarthritis, the patient sample is analyzed, and the results of the analysis may be utilized to diagnose osteoarthritis. Also osteoarthritis diagnostic method can be developed from studies used to generate prognostic methods in which populations are stratified into subpopulations having different progressions of osteoarthritis. In another embodiment, prognostic results may be gathered, a patient's risk factors for developing osteoarthritis (e.g., age, weight, race, diet) analyzed, and a patient sample may be ordered based on a determined predisposition to osteoarthritis.

[0085] The nucleic acid sample typically is isolated from a biological sample obtained from a subject. For example, nucleic acid can be isolated from blood, saliva, sputum, urine, cell scrapings, and biopsy tissue. The nucleic acid sample can be isolated from a biological sample using standard techniques, such as the technique described in Example 2. As used herein, the term "subject" refers primarily to humans but also refers to other mammals such as dogs, cats, and ungulates (e.g., cattle, sheep, and swine). Subjects also include avians (e.g., chickens and turkeys), reptiles, and fish (e.g., salmon), as embodiments described herein can be adapted to nucleic acid samples isolated from any of these organisms. The nucleic acid sample may be isolated from the subject and then directly utilized in a method for determining the presence of a polymorphic variant, or alternatively, the sample may be isolated and then stored (e.g., frozen) for a period of time before being subjected to analysis.

[0086] The presence or absence of a polymorphic variant is determined using one or both chromosomal complements represented in the nucleic acid sample. Determining the presence or absence of a polymorphic variant in both chromosomal complements represented in a nucleic acid sample from a subject having a copy of each chromosome is useful for determining the zygosity of an individual for the polymorphic variant (i.e., whether the individual is homozygous or heterozygous for the polymorphic variant). Any oligonucleotide-based diagnostic may be utilized to determine whether a sample includes the presence or absence of a polymorphic variant in a sample. For example, primer extension methods, ligase sequence determination methods (e.g., U.S. Pat. Nos. 5,679,524 and 5,952,174, and WO

01/27326), mismatch sequence determination methods (*e.g.*, U.S. Pat. Nos. 5,851,770; 5,958,692; 6,110,684; and 6,183,958), microarray sequence determination methods, restriction fragment length polymorphism (RFLP), single strand conformation polymorphism detection (SSCP) (*e.g.*, U.S. Pat. Nos. 5,891,625 and 6,013,499), PCR-based assays (*e.g.*, TAQMAN[®] PCR System (Applied Biosystems)), and nucleotide sequencing methods may be used.

[0087] Oligonucleotide extension methods typically involve providing a pair of oligonucleotide primers in a polymerase chain reaction (PCR) or in other nucleic acid amplification methods for the purpose of amplifying a region from the nucleic acid sample that comprises the polymorphic variation. One oligonucleotide primer is complementary to a region 3' of the polymorphism and the other is complementary to a region 5' of the polymorphism. A PCR primer pair may be used in methods disclosed in U.S. Pat. Nos. 4,683,195; 4,683,202; 4,965,188; 5,656,493; 5,998,143; 6,140,054; WO 01/27327; and WO 01/27329 for example. PCR primer pairs may also be used in any commercially available machines that perform PCR, such as any of the GENEAMP[®] Systems available from Applied Biosystems. Also, those of ordinary skill in the art will be able to design oligonucleotide primers based upon a *TM7SF3* nucleotide sequence using knowledge available in the art.

[0088] Also provided is an extension oligonucleotide that hybridizes to the amplified fragment adjacent to the polymorphic variation. As used herein, the term "adjacent" refers to the 3' end of the extension oligonucleotide being often 1 nucleotide from the 5' end of the polymorphic site, and sometimes 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides from the 5' end of the polymorphic site, in the nucleic acid when the extension oligonucleotide is hybridized to the nucleic acid. The extension oligonucleotide then is extended by one or more nucleotides, and the number and/or type of nucleotides that are added to the extension oligonucleotide determine whether the polymorphic variant is present. Oligonucleotide extension methods are disclosed, for example, in U.S. Pat. Nos. 4,656,127; 4,851,331; 5,679,524; 5,834,189; 5,876,934; 5,908,755; 5,912,118; 5,976,802; 5,981,186; 6,004,744; 6,013,431; 6,017,702; 6,046,005; 6,087,095; 6,210,891; and WO 01/20039. Oligonucleotide extension methods using mass spectrometry are described, for example, in U.S. Pat. Nos. 5,547,835; 5,605,798; 5,691,141; 5,849,542; 5,869,242; 5,928,906; 6,043,031; and 6,194,144, and a method often utilized is described herein in Example 2.

[0089] A microarray can be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A microarray may include any oligonucleotides described herein, and methods for making and using oligonucleotide microarrays suitable for diagnostic use are disclosed in U.S. Pat. Nos. 5,492,806; 5,525,464; 5,589,330; 5,695,940; 5,849,483; 6,018,041; 6,045,996; 6,136,541; 6,142,681; 6,156,501; 6,197,506; 6,223,127; 6,225,625; 6,229,911; 6,239,273; WO 00/52625; WO 01/25485; and WO 01/29259. The microarray typically comprises a solid support and the oligonucleotides may be linked to this solid support by covalent bonds or by non-covalent interactions.

The oligonucleotides may also be linked to the solid support directly or by a spacer molecule. A microarray may comprise one or more oligonucleotides complementary to a polymorphic site set forth herein.

[0090] A kit also may be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A kit often comprises one or more pairs of oligonucleotide primers useful for amplifying a fragment of a nucleotide sequence of SEQ ID NO: 1-2 or a substantially identical sequence thereof, where the fragment includes a polymorphic site. The kit sometimes comprises a polymerizing agent, for example, a thermostable nucleic acid polymerase such as one disclosed in U.S. Pat. Nos. 4,889,818 or 6,077,664. Also, the kit often comprises an elongation oligonucleotide that hybridizes to a *TM7SF3* nucleotide sequence in a nucleic acid sample adjacent to the polymorphic site. Where the kit includes an elongation oligonucleotide, it also often comprises chain elongating nucleotides, such as dATP, dTTP, dGTP, dCTP, and dITP, including analogs of dATP, dTTP, dGTP, dCTP and dITP, provided that such analogs are substrates for a thermostable nucleic acid polymerase and can be incorporated into a nucleic acid chain elongated from the extension oligonucleotide. Along with chain elongating nucleotides would be one or more chain terminating nucleotides such as ddATP, ddTTP, ddGTP, ddCTP, and the like. In an embodiment, the kit comprises one or more oligonucleotide primer pairs, a polymerizing agent, chain elongating nucleotides, at least one elongation oligonucleotide, and one or more chain terminating nucleotides. Kits optionally include buffers, vials, microtiter plates, and instructions for use.

[0091] An individual identified as being at risk of osteoarthritis may be heterozygous or homozygous with respect to the allele associated with a higher risk of osteoarthritis. A subject homozygous for an allele associated with an increased risk of osteoarthritis is at a comparatively high risk of osteoarthritis, a subject heterozygous for an allele associated with an increased risk of osteoarthritis is at a comparatively intermediate risk of osteoarthritis, and a subject homozygous for an allele associated with a decreased risk of osteoarthritis is at a comparatively low risk of osteoarthritis. A genotype may be assessed for a complementary strand, such that the complementary nucleotide at a particular position is detected.

[0092] Also featured are methods for determining risk of osteoarthritis and/or identifying a subject at risk of osteoarthritis by contacting a polypeptide or protein encoded by a *TM7SF3* nucleotide sequence from a subject with an antibody that specifically binds to an epitope associated with increased risk of osteoarthritis in the polypeptide.

Applications of Prognostic and Diagnostic Results to Pharmacogenomic Methods

[0093] Pharmacogenomics is a discipline that involves tailoring a treatment for a subject according to the subject's genotype as a particular treatment regimen may exert a differential effect depending upon

the subject's genotype. For example, based upon the outcome of a prognostic test described herein, a clinician or physician may target pertinent information and preventative or therapeutic treatments to a subject who would be benefited by the information or treatment and avoid directing such information and treatments to a subject who would not be benefited (*e.g.*, the treatment has no therapeutic effect and/or the subject experiences adverse side effects).

[0094] The following is an example of a pharmacogenomic embodiment. A particular treatment regimen can exert a differential effect depending upon the subject's genotype. Where a candidate therapeutic exhibits a significant interaction with a major allele and a comparatively weak interaction with a minor allele (*e.g.*, an order of magnitude or greater difference in the interaction), such a therapeutic typically would not be administered to a subject genotyped as being homozygous for the minor allele, and sometimes not administered to a subject genotyped as being heterozygous for the minor allele. In another example, where a candidate therapeutic is not significantly toxic when administered to subjects who are homozygous for a major allele but is comparatively toxic when administered to subjects heterozygous or homozygous for a minor allele, the candidate therapeutic is not typically administered to subjects who are genotyped as being heterozygous or homozygous with respect to the minor allele.

[0095] The methods described herein are applicable to pharmacogenomic methods for preventing, alleviating or treating osteoarthritis. For example, a nucleic acid sample from an individual may be subjected to a prognostic test described herein. Where one or more polymorphic variations associated with increased risk of osteoarthritis are identified in a subject, information for preventing or treating osteoarthritis and/or one or more osteoarthritis treatment regimens then may be prescribed to that subject.

[0096] In certain embodiments, a treatment or preventative regimen is specifically prescribed and/or administered to individuals who will most benefit from it based upon their risk of developing osteoarthritis assessed by the methods described herein. Thus, provided are methods for identifying a subject predisposed to osteoarthritis and then prescribing a therapeutic or preventative regimen to individuals identified as having a predisposition. Thus, certain embodiments are directed to a method for reducing osteoarthritis in a subject, which comprises: detecting the presence or absence of a polymorphic variant associated with osteoarthritis in a nucleotide sequence in a nucleic acid sample from a subject, where the nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) a nucleotide sequence of SEQ ID NO: 1-2; (b) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-2; and (d) a fragment of a polynucleotide sequence of (a), (b), or (c); and prescribing or administering a treatment regimen to a subject from whom the sample originated where the presence of a polymorphic variation associated with osteoarthritis is

detected in the nucleotide sequence. In these methods, predisposition results may be utilized in combination with other test results to diagnose osteoarthritis.

[0097] Certain preventative treatments often are prescribed to subjects having a predisposition to osteoarthritis and where the subject is diagnosed with osteoarthritis or is diagnosed as having symptoms indicative of an early stage of osteoarthritis. The treatment sometimes is preventative (*e.g.*, is prescribed or administered to reduce the probability that osteoarthritis arises or progresses), sometimes is therapeutic, and sometimes delays, alleviates or halts the progression of osteoarthritis. Any known preventative or therapeutic treatment for alleviating or preventing the occurrence of osteoarthritis is prescribed and/or administered. For example, the treatment often is directed to decreasing pain and improving joint movement. Examples of OA treatments include exercises to keep joints flexible and improve muscle strength. Different medications to control pain, including corticosteroids and nonsteroidal anti-inflammatory drugs (NSAIDs, *e.g.*, Voltaren); cyclooxygenase-2 (COX-2) inhibitors (*e.g.*, Celebrex, Vioxx, Mobic, and Bextra); monoclonal antibodies (*e.g.*, Remicade); tumor necrosis factor inhibitors (*e.g.*, Enbrel); or injections of glucocorticoids, hyaluronic acid or chondrotin sulfate into joints that are inflamed and not responsive to NSAIDs. Orally administered chondroitin sulfate also may be used as a therapeutic, as it may increase hyaluronic acid levels and viscosity of synovial fluid, and decrease collagenase levels in synovial fluid. Also, glucosamine can serve as an OA therapeutic as delivering it into joints may inhibit enzymes involved in cartilage degradation and enhance the production of hyaluronic acid. For mild pain without inflammation, acetaminophen may be used. Other treatments include: heat/cold therapy for temporary pain relief; joint protection to prevent strain or stress on painful joints; surgery to relieve chronic pain in damaged joints; and weight control to prevent extra stress on weight-bearing joints.

[0098] As therapeutic approaches for treating osteoarthritis continue to evolve and improve, the goal of treatments for osteoarthritis related disorders is to intervene even before clinical signs first manifest. Thus, genetic markers associated with susceptibility to osteoarthritis prove useful for early diagnosis, prevention and treatment of osteoarthritis.

[0099] As osteoarthritis preventative and treatment information can be specifically targeted to subjects in need thereof (*e.g.*, those at risk of developing osteoarthritis or those in an early stage of osteoarthritis), provided herein is a method for preventing or reducing the risk of developing osteoarthritis in a subject, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying a subject with a predisposition to osteoarthritis, whereby the presence of the polymorphic variation is indicative of a predisposition to osteoarthritis in the subject; and (c) if such a predisposition is identified, providing the subject with information about methods or products to prevent or reduce osteoarthritis or to delay the onset of osteoarthritis. Also provided is a

method of targeting information or advertising to a subpopulation of a human population based on the subpopulation being genetically predisposed to a disease or condition, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying the subpopulation of subjects in which the polymorphic variation is associated with osteoarthritis; and (c) providing information only to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition.

[0100] Pharmacogenomics methods also may be used to analyze and predict a response to osteoarthritis treatment or a drug. For example, if pharmacogenomics analysis indicates a likelihood that an individual will respond positively to osteoarthritis treatment with a particular drug, the drug may be administered to the individual. Conversely, if the analysis indicates that an individual is likely to respond negatively to treatment with a particular drug, an alternative course of treatment may be prescribed. A negative response may be defined as either the absence of an efficacious response or the presence of toxic side effects. The response to a therapeutic treatment can be predicted in a background study in which subjects in any of the following populations are genotyped: a population that responds favorably to a treatment regimen, a population that does not respond significantly to a treatment regimen, and a population that responds adversely to a treatment regimen (*e.g.*, exhibits one or more side effects). These populations are provided as examples and other populations and subpopulations may be analyzed. Based upon the results of these analyses, a subject is genotyped to predict whether he or she will respond favorably to a treatment regimen, not respond significantly to a treatment regimen, or respond adversely to a treatment regimen.

[0101] The tests described herein also are applicable to clinical drug trials. One or more polymorphic variants indicative of response to an agent for treating osteoarthritis or to side effects to an agent for treating osteoarthritis may be identified using the methods described herein. Thereafter, potential participants in clinical trials of such an agent may be screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment may be measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems.

[0102] Thus, another embodiment is a method of selecting an individual for inclusion in a clinical trial of a treatment or drug comprising the steps of: (a) obtaining a nucleic acid sample from an individual; (b) determining the identity of a polymorphic variation which is associated with a positive response to the treatment or the drug, or at least one polymorphic variation which is associated with a negative response to the treatment or the drug in the nucleic acid sample, and (c) including the individual in the clinical trial if the nucleic acid sample contains said polymorphic variation associated with a

positive response to the treatment or the drug or if the nucleic acid sample lacks said polymorphic variation associated with a negative response to the treatment or the drug. In addition, the methods described herein for selecting an individual for inclusion in a clinical trial of a treatment or drug encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination. The polymorphic variation may be in a sequence selected individually or in any combination from the group consisting of (i) a nucleotide sequence of SEQ ID NO: 1-2; (ii) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2; (iii) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-2; and (iv) a fragment of a polynucleotide sequence of (i), (ii), or (iii) comprising the polymorphic site. The including step (c) optionally comprises administering the drug or the treatment to the individual if the nucleic acid sample contains the polymorphic variation associated with a positive response to the treatment or the drug and the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug.

[0103] Also provided herein is a method of partnering between a diagnostic/prognostic testing provider and a provider of a consumable product, which comprises: (a) the diagnostic/prognostic testing provider detects the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) the diagnostic/prognostic testing provider identifies the subpopulation of subjects in which the polymorphic variation is associated with osteoarthritis; (c) the diagnostic/prognostic testing provider forwards information to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition; and (d) the provider of a consumable product forwards to the diagnostic test provider a fee every time the diagnostic/prognostic test provider forwards information to the subject as set forth in step (c) above.

Compositions Comprising Osteoarthritis-Directed Molecules

[0104] Featured herein is a composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and one or more molecules specifically directed and targeted to a nucleic acid comprising a *TM7SF3* nucleotide sequence or amino acid sequence. Such directed molecules include, but are not limited to, a compound that binds to a *TM7SF3* nucleotide sequence or amino acid sequence referenced herein; a RNAi or siRNA molecule having a strand complementary or substantially complementary to a *TM7SF3* nucleotide sequence (e.g., hybridizes to a *TM7SF3* nucleotide sequence under conditions of high stringency); an antisense nucleic acid complementary or substantially complementary to an RNA encoded by a *TM7SF3* nucleotide sequence (e.g., hybridizes to a *TM7SF3*

nucleotide sequence under conditions of high stringency); a ribozyme that hybridizes to a *TM7SF3* nucleotide sequence (e.g., hybridizes to a *TM7SF3* nucleotide sequence under conditions of high stringency); a nucleic acid aptamer that specifically binds a polypeptide encoded by *TM7SF3* nucleotide sequence; and an antibody that specifically binds to a polypeptide encoded by *TM7SF3* nucleotide sequence or binds to a nucleic acid having such a nucleotide sequence. In specific embodiments, the osteoarthritis directed molecule interacts with a nucleic acid or polypeptide variant associated with osteoarthritis, such as variants referenced herein. In other embodiments, the osteoarthritis directed molecule interacts with a polypeptide involved in a signal pathway of a polypeptide encoded by a *TM7SF3* nucleotide sequence, or a nucleic acid comprising such a nucleotide sequence.

[0105] Compositions sometimes include an adjuvant known to stimulate an immune response, and in certain embodiments, an adjuvant that stimulates a T-cell lymphocyte response. Adjuvants are known, including but not limited to an aluminum adjuvant (e.g., aluminum hydroxide); a cytokine adjuvant or adjuvant that stimulates a cytokine response (e.g., interleukin (IL)-12 and/or gamma-interferon cytokines); a Freund-type mineral oil adjuvant emulsion (e.g., Freund's complete or incomplete adjuvant); a synthetic lipoid compound; a copolymer adjuvant (e.g., TitreMax); a saponin; Quil A; a liposome; an oil-in-water emulsion (e.g., an emulsion stabilized by Tween 80 and pluronic polyoxyethylene/polyoxypropylene block copolymer (Syntex Adjuvant Formulation); TitreMax; detoxified endotoxin (MPL) and mycobacterial cell wall components (TDW, CWS) in 2% squalene (Ribi Adjuvant System)); a muramyl dipeptide; an immune-stimulating complex (ISCOM, e.g., an Ag-modified saponin/cholesterol micelle that forms stable cage-like structure); an aqueous phase adjuvant that does not have a depot effect (e.g., Gerbu adjuvant); a carbohydrate polymer (e.g., AdjuPrime); L-tyrosine; a manide-oleate compound (e.g., Montanide); an ethylene-vinyl acetate copolymer (e.g., Elvax 40W1,2); or lipid A, for example. Such compositions are useful for generating an immune response against osteoarthritis directed molecule (e.g., an HLA-binding subsequence within a polypeptide encoded by a *TM7SF3* nucleotide sequence). In such methods, a peptide having an amino acid subsequence of a polypeptide encoded by a *TM7SF3* nucleotide sequence is delivered to a subject, where the subsequence binds to an HLA molecule and induces a CTL lymphocyte response. The peptide sometimes is delivered to the subject as an isolated peptide or as a minigene in a plasmid that encodes the peptide. Methods for identifying HLA-binding subsequences in such polypeptides are known (see e.g., publication WO02/20616 and PCT application US98/01373 for methods of identifying such sequences).

[0106] The cell may be in a group of cells cultured *in vitro* or in a tissue maintained *in vitro* or present in an animal *in vivo* (e.g., a rat, mouse, ape or human). In certain embodiments, a composition comprises a component from a cell such as a nucleic acid molecule (e.g., genomic DNA), a protein mixture or isolated protein, for example. The aforementioned compositions have utility in diagnostic,

prognostic and pharmacogenomic methods described previously and in therapeutics described hereafter. Certain osteoarthritis directed molecules are described in greater detail below.

Compounds

[0107] Compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive (see, e.g., Zuckermann et al., *J. Med. Chem.* 37: 2678-85 (1994)); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; "one-bead one-compound" library methods; and synthetic library methods using affinity chromatography selection. Biological library and peptoid library approaches are typically limited to peptide libraries, while the other approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, *Anticancer Drug Des.* 12: 145, (1997)). Examples of methods for synthesizing molecular libraries are described, for example, in DeWitt et al., *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909 (1993); Erb et al., *Proc. Natl. Acad. Sci. USA* 91: 11422 (1994); Zuckermann et al., *J. Med. Chem.* 37: 2678 (1994); Cho et al., *Science* 261: 1303 (1993); Carrell et al., *Angew. Chem. Int. Ed. Engl.* 33: 2059 (1994); Carrell et al., *Angew. Chem. Int. Ed. Engl.* 33: 2061 (1994); and in Gallop et al., *J. Med. Chem.* 37: 1233 (1994).

[0108] Libraries of compounds may be presented in solution (e.g., Houghten, *Biotechniques* 13: 412-421 (1992)), or on beads (Lam, *Nature* 354: 82-84 (1991)), chips (Fodor, *Nature* 364: 555-556 (1993)), bacteria or spores (Ladner, United States Patent No. 5,223,409), plasmids (Cull et al., *Proc. Natl. Acad. Sci. USA* 89: 1865-1869 (1992)) or on phage (Scott and Smith, *Science* 249: 386-390 (1990); Devlin, *Science* 249: 404-406 (1990); Cwirla et al., *Proc. Natl. Acad. Sci.* 87: 6378-6382 (1990); Felici, *J. Mol. Biol.* 222: 301-310 (1991); Ladner supra.).

[0109] A compound sometimes alters expression and sometimes alters activity of a polypeptide target and may be a small molecule. Small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

Antisense Nucleic Acid Molecules, Ribozymes, RNAi, siRNA and Modified Nucleic Acid Molecules

[0110] An “antisense” nucleic acid refers to a nucleotide sequence complementary to a “sense” nucleic acid encoding a polypeptide, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire coding strand, or to a portion thereof or a substantially identical sequence thereof. In another embodiment, the antisense nucleic acid molecule is antisense to a “noncoding region” of the coding strand of a nucleotide sequence (e.g., 5’ and 3’ untranslated regions in SEQ ID NO: 1).

[0111] An antisense nucleic acid can be designed such that it is complementary to the entire coding region of an mRNA encoded by a nucleotide sequence (e.g., SEQ ID NO: 1), and often the antisense nucleic acid is an oligonucleotide antisense to only a portion of a coding or noncoding region of the mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of the mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length. The antisense nucleic acids, which include the ribozymes described hereafter, can be designed to target a *TM7SF3* nucleotide sequence, often a variant associated with osteoarthritis, or a substantially identical sequence thereof. Among the variants, minor alleles and major alleles can be targeted, and those associated with a higher risk of osteoarthritis are often designed, tested, and administered to subjects.

[0112] An antisense nucleic acid can be constructed using chemical synthesis and enzymatic ligation reactions using standard procedures. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0113] When utilized as therapeutics, antisense nucleic acids typically are administered to a subject (e.g., by direct injection at a tissue site) or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a polypeptide and thereby inhibit expression of the polypeptide, for example, by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then are administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, for example, by linking antisense nucleic acid molecules

to peptides or antibodies which bind to cell surface receptors or antigens. Antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. Sufficient intracellular concentrations of antisense molecules are achieved by incorporating a strong promoter, such as a pol II or pol III promoter, in the vector construct.

[0114] Antisense nucleic acid molecules sometimes are alpha-anomeric nucleic acid molecules. An alpha-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual beta-units, the strands run parallel to each other (Gaultier et al., *Nucleic Acids. Res.* 15: 6625-6641 (1987)). Antisense nucleic acid molecules can also comprise a 2'-O-methylribonucleotide (Inoue et al., *Nucleic Acids Res.* 15: 6131-6148 (1987)) or a chimeric RNA-DNA analogue (Inoue et al., *FEBS Lett.* 215: 327-330 (1987)). Antisense nucleic acids sometimes are composed of DNA or PNA or any other nucleic acid derivatives described previously.

[0115] In another embodiment, an antisense nucleic acid is a ribozyme. A ribozyme having specificity for a *TM7SF3* nucleotide sequence can include one or more sequences complementary to such a nucleotide sequence, and a sequence having a known catalytic region responsible for mRNA cleavage (see e.g., U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach, *Nature* 334: 585-591 (1988)). For example, a derivative of a Tetrahymena L-19 IVS RNA is sometimes utilized in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a mRNA (see e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742). Also, target mRNA sequences can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (see e.g., Bartel & Szostak, *Science* 261: 1411-1418 (1993)).

[0116] Osteoarthritis directed molecules include in certain embodiments nucleic acids that can form triple helix structures with a *TM7SF3* nucleotide sequence, or a substantially identical sequence thereof, especially one that includes a regulatory region that controls expression of a polypeptide. Gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of a nucleotide sequence referenced herein or a substantially identical sequence (e.g., promoter and/or enhancers) to form triple helical structures that prevent transcription of a gene in target cells (see e.g., Helene, *Anticancer Drug Des.* 6(6): 569-84 (1991); Helene et al., *Ann. N.Y. Acad. Sci.* 660: 27-36 (1992); and Maher, *Bioassays* 14(12): 807-15 (1992). Potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[0117] Osteoarthritis directed molecules include RNAi and siRNA nucleic acids. Gene expression may be inhibited by the introduction of double-stranded RNA (dsRNA), which induces potent and specific gene silencing, a phenomenon called RNA interference or RNAi. See, e.g., Fire et al., US Patent

Number 6,506,559; Tuschl et al. PCT International Publication No. WO 01/75164; Kay et al. PCT International Publication No. WO 03/010180A1; or Bosher JM, Labouesse, Nat Cell Biol 2000 Feb;2(2):E31-6. This process has been improved by decreasing the size of the double-stranded RNA to 20-24 base pairs (to create small-interfering RNAs or siRNAs) that “switched off” genes in mammalian cells without initiating an acute phase response, i.e., a host defense mechanism that often results in cell death (see, e.g., Caplen et al. Proc Natl Acad Sci U S A. 2001 Aug 14;98(17):9742-7 and Elbashir et al. Methods 2002 Feb;26(2):199-213). There is increasing evidence of post-transcriptional gene silencing by RNA interference (RNAi) for inhibiting targeted expression in mammalian cells at the mRNA level, in human cells. There is additional evidence of effective methods for inhibiting the proliferation and migration of tumor cells in human patients, and for inhibiting metastatic cancer development (see, e.g., U.S. Patent Application No. US2001000993183; Caplen et al. Proc Natl Acad Sci U S A; and Abderrahmani et al. Mol Cell Biol 2001 Nov21(21):7256-67).

[0118] An “siRNA” or “RNAi” refers to a nucleic acid that forms a double stranded RNA and has the ability to reduce or inhibit expression of a gene or target gene when the siRNA is delivered to or expressed in the same cell as the gene or target gene. “siRNA” refers to short double-stranded RNA formed by the complementary strands. Complementary portions of the siRNA that hybridize to form the double stranded molecule often have substantial or complete identity to the target molecule sequence. In one embodiment, an siRNA refers to a nucleic acid that has substantial or complete identity to a target gene and forms a double stranded siRNA.

[0119] When designing the siRNA molecules, the targeted region often is selected from a given DNA sequence beginning 50 to 100 nucleotides downstream of the start codon. See, e.g., Elbashir et al., Methods 26:199-213 (2002). Initially, 5’ or 3’ UTRs and regions nearby the start codon were avoided assuming that UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. Sometimes regions of the target 23 nucleotides in length conforming to the sequence motif AA(N19)TT (N, an nucleotide), and regions with approximately 30% to 70% G/C-content (often about 50% G/C-content) often are selected. If no suitable sequences are found, the search often is extended using the motif NA(N21). The sequence of the sense siRNA sometimes corresponds to (N19) TT or N21 (position 3 to 23 of the 23-nt motif), respectively. In the latter case, the 3’ end of the sense siRNA often is converted to TT. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and antisense 3’ overhangs. The antisense siRNA is synthesized as the complement to position 1 to 21 of the 23-nt motif. Because position 1 of the 23-nt motif is not recognized sequence-specifically by the antisense siRNA, the 3’-most nucleotide residue of the antisense siRNA can be chosen deliberately. However, the penultimate nucleotide of the antisense siRNA (complementary to position 2 of the 23-nt motif) often is complementary to the targeted sequence. For simplifying chemical synthesis, TT often is

utilized. siRNAs corresponding to the target motif NAR(N17)YNN, where R is purine (A,G) and Y is pyrimidine (C,U), often are selected. Respective 21 nucleotide sense and antisense siRNAs often begin with a purine nucleotide and can also be expressed from pol III expression vectors without a change in targeting site. Expression of RNAs from pol III promoters often is efficient when the first transcribed nucleotide is a purine.

[0120] The sequence of the siRNA can correspond to the full length target gene, or a subsequence thereof. Often, the siRNA is about 15 to about 50 nucleotides in length (e.g., each complementary sequence of the double stranded siRNA is 15-50 nucleotides in length, and the double stranded siRNA is about 15-50 base pairs in length, sometimes about 20-30 nucleotides in length or about 20-25 nucleotides in length, e.g., 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. The siRNA sometimes is about 21 nucleotides in length. Methods of using siRNA are well known in the art, and specific siRNA molecules may be purchased from a number of companies including Dharmacon Research, Inc.

[0121] Antisense, ribozyme, RNAi and siRNA nucleic acids can be altered to form modified nucleic acid molecules. The nucleic acids can be altered at base moieties, sugar moieties or phosphate backbone moieties to improve stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup et al., *Bioorganic & Medicinal Chemistry* 4 (1): 5-23 (1996)). As used herein, the terms “peptide nucleic acid” or “PNA” refers to a nucleic acid mimic such as a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. Synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described, for example, in Hyrup et al., (1996) *supra* and Perry-O’Keefe et al., *Proc. Natl. Acad. Sci.* 93: 14670-675 (1996).

[0122] PNA nucleic acids can be used in prognostic, diagnostic, and therapeutic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNA nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as “artificial restriction enzymes” when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup et al., (1996) *supra*; Perry-O’Keefe *supra*).

[0123] In other embodiments, oligonucleotides may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across cell membranes (see e.g., Letsinger et al., *Proc. Natl. Acad. Sci. USA* 86: 6553-6556 (1989); Lemaitre et al., *Proc. Natl. Acad. Sci. USA* 84: 648-652 (1987); PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-

triggered cleavage agents (See, e.g., Krol et al., Bio-Techniques 6: 958-976 (1988)) or intercalating agents. (See, e.g., Zon, Pharm. Res. 5: 539-549 (1988)). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[0124] Also included herein are molecular beacon oligonucleotide primer and probe molecules having one or more regions complementary to a *TM7SF3* nucleotide sequence, or a substantially identical sequence thereof, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantifying the presence of the nucleic acid in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi et al., U.S. Patent No. 5,854,033; Nazarenko et al., U.S. Patent No. 5,866,336, and Livak et al., U.S. Patent 5,876,930.

Antibodies

[0125] The term “antibody” as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin. An antibody sometimes is a polyclonal, monoclonal, recombinant (e.g., a chimeric or humanized), fully human, non-human (e.g., murine), or a single chain antibody. An antibody may have effector function and can fix complement, and is sometimes coupled to a toxin or imaging agent.

[0126] A full-length polypeptide or antigenic peptide fragment encoded by a nucleotide sequence referenced herein can be used as an immunogen or can be used to identify antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. An antigenic peptide often includes at least 8 amino acid residues of the amino acid sequences encoded by a nucleotide sequence referenced herein, or substantially identical sequence thereof, and encompasses an epitope. Antigenic peptides sometimes include 10 or more amino acids, 15 or more amino acids, 20 or more amino acids, or 30 or more amino acids. Hydrophilic and hydrophobic fragments of polypeptides sometimes are used as immunogens.

[0127] Epitopes encompassed by the antigenic peptide are regions located on the surface of the polypeptide (e.g., hydrophilic regions) as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human polypeptide sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the polypeptide and are thus likely to constitute surface residues useful for targeting antibody production. The antibody may bind an epitope on any domain or region on polypeptides described herein.

[0128] Also, chimeric, humanized, and completely human antibodies are useful for applications which include repeated administration to subjects. Chimeric and humanized monoclonal antibodies,

comprising both human and non-human portions, can be made using standard recombinant DNA techniques. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al International Application No. PCT/US86/02269; Akira, et al European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al European Patent Application 173,494; Neuberger et al PCT International Publication No. WO 86/01533; Cabilly et al U.S. Patent No. 4,816,567; Cabilly et al European Patent Application 125,023; Better et al., Science 240: 1041-1043 (1988); Liu et al., Proc. Natl. Acad. Sci. USA 84: 3439-3443 (1987); Liu et al., J. Immunol. 139: 3521-3526 (1987); Sun et al., Proc. Natl. Acad. Sci. USA 84: 214-218 (1987); Nishimura et al., Canc. Res. 47: 999-1005 (1987); Wood et al., Nature 314: 446-449 (1985); and Shaw et al., J. Natl. Cancer Inst. 80: 1553-1559 (1988); Morrison, S. L., Science 229: 1202-1207 (1985); Oi et al., BioTechniques 4: 214 (1986); Winter U.S. Patent 5,225,539; Jones et al., Nature 321: 552-525 (1986); Verhoeyan et al., Science 239: 1534; and Beidler et al., J. Immunol. 141: 4053-4060 (1988).

[0129] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice that are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. See, for example, Lonberg and Huszar, Int. Rev. Immunol. 13: 65-93 (1995); and U.S. Patent Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, CA) and Medarex, Inc. (Princeton, NJ), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above. Completely human antibodies that recognize a selected epitope also can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody (e.g., a murine antibody) is used to guide the selection of a completely human antibody recognizing the same epitope. This technology is described for example by Jespers et al., Bio/Technology 12: 899-903 (1994).

[0130] An antibody can be a single chain antibody. A single chain antibody (scFV) can be engineered (see, e.g., Colcher et al., Ann. N Y Acad. Sci. 880: 263-80 (1999); and Reiter, Clin. Cancer Res. 2: 245-52 (1996)). Single chain antibodies can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target polypeptide.

[0131] Antibodies also may be selected or modified so that they exhibit reduced or no ability to bind an Fc receptor. For example, an antibody may be an isotype or subtype, fragment or other mutant, which does not support binding to an Fc receptor (e.g., it has a mutagenized or deleted Fc receptor binding region).

[0132] Also, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium

bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1 dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thiotepa chlorambucil, melphalan, carmustine (BCNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

[0133] Antibody conjugates can be used for modifying a given biological response. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, gamma-interferon, alpha-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Also, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980, for example.

[0134] An antibody (e.g., monoclonal antibody) can be used to isolate target polypeptides by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an antibody can be used to detect a target polypeptide (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. Antibodies can be used diagnostically to monitor polypeptide levels in tissue as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H . Also, an antibody can be utilized as a test

molecule for determining whether it can treat osteoarthritis, and as a therapeutic for administration to a subject for treating osteoarthritis.

[0135] An antibody can be made by immunizing with a purified antigen, or a fragment thereof, e.g., a fragment described herein, a membrane associated antigen, tissues, e.g., crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions.

[0136] Included herein are antibodies which bind only a native polypeptide, only denatured or otherwise non-native polypeptide, or which bind both, as well as those having linear or conformational epitopes. Conformational epitopes sometimes can be identified by selecting antibodies that bind to native but not denatured polypeptide. Also featured are antibodies that specifically bind to a polypeptide variant associated with osteoarthritis.

Methods for Identifying Candidate Therapeutics for Treating Osteoarthritis

[0137] Current therapies for the treatment of osteoarthritis have limited efficacy, limited tolerability and significant mechanism-based side effects, and few of the available therapies adequately address underlying defects. Current therapeutic approaches were largely developed in the absence of defined molecular targets or even a solid understanding of disease pathogenesis. Therefore, provided are methods of identifying candidate therapeutics that target biochemical pathways related to the development of osteoarthritis.

[0138] Thus, featured herein are methods for identifying a candidate therapeutic for treating osteoarthritis. The methods comprise contacting a test molecule with a target molecule in a system. A “target molecule” as used herein refers to a *TM7SF3* nucleic acid, a substantially identical nucleic acid thereof, or a fragment thereof, and an encoded polypeptide of the foregoing. The methods also comprise determining the presence or absence of an interaction between the test molecule and the target molecule, where the presence of an interaction between the test molecule and the nucleic acid or polypeptide identifies the test molecule as a candidate osteoarthritis therapeutic. The interaction between the test molecule and the target molecule may be quantified.

[0139] Test molecules and candidate therapeutics include, but are not limited to, compounds, antisense nucleic acids, siRNA molecules, ribozymes, polypeptides or proteins encoded by a *TM7SF3* nucleotide sequence, or a substantially identical sequence or fragment thereof, and immunotherapeutics (e.g., antibodies and HLA-presented polypeptide fragments). A test molecule or candidate therapeutic may act as a modulator of target molecule concentration or target molecule function in a system. A “modulator” may agonize (i.e., up-regulates) or antagonize (i.e., down-regulates) a target molecule concentration partially or completely in a system by affecting such cellular functions as DNA replication and/or DNA processing (e.g., DNA methylation or DNA repair), RNA transcription and/or RNA processing (e.g., removal of intronic sequences and/or translocation of spliced mRNA from the nucleus),

polypeptide production (e.g., translation of the polypeptide from mRNA), and/or polypeptide post-translational modification (e.g., glycosylation, phosphorylation, and proteolysis of pro-polypeptides). A modulator may also agonize or antagonize a biological function of a target molecule partially or completely, where the function may include adopting a certain structural conformation, interacting with one or more binding partners, ligand binding, catalysis (e.g., phosphorylation, dephosphorylation, hydrolysis, methylation, and isomerization), and an effect upon a cellular event (e.g., effecting progression of osteoarthritis).

[0140] As used herein, the term “system” refers to a cell free *in vitro* environment and a cell-based environment such as a collection of cells, a tissue, an organ, or an organism. A system is “contacted” with a test molecule in a variety of manners, including adding molecules in solution and allowing them to interact with one another by diffusion, cell injection, and any administration routes in an animal. As used herein, the term “interaction” refers to an effect of a test molecule on test molecule, where the effect sometimes is binding between the test molecule and the target molecule, and sometimes is an observable change in cells, tissue, or organism.

[0141] There are many standard methods for detecting the presence or absence of interaction between a test molecule and a target molecule. For example, titrametric, acidimetric, radiometric, NMR, monolayer, polarographic, spectrophotometric, fluorescent, and ESR assays probative of a target molecule interaction may be utilized.

[0142] Test molecule/target molecule interactions can be detected and/or quantified using assays known in the art. For example, an interaction can be determined by labeling the test molecule and/or the target molecule, where the label is covalently or non-covalently attached to the test molecule or target molecule. The label is sometimes a radioactive molecule such as ^{125}I , ^{131}I , ^{35}S or ^3H , which can be detected by direct counting of radioemission or by scintillation counting. Also, enzymatic labels such as horseradish peroxidase, alkaline phosphatase, or luciferase may be utilized where the enzymatic label can be detected by determining conversion of an appropriate substrate to product. In addition, presence or absence of an interaction can be determined without labeling. For example, a microphysiometer (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indication of an interaction between a test molecule and target molecule (McConnell, H. M. *et al.*, *Science* 257: 1906-1912 (1992)).

[0143] In cell-based systems, cells typically include a *TM7SF3* nucleic acid, an encoded polypeptide, or substantially identical nucleic acid or polypeptide thereof, and are often of mammalian origin, although the cell can be of any origin. Whole cells, cell homogenates, and cell fractions (e.g., cell membrane fractions) can be subjected to analysis. Where interactions between a test molecule with a target polypeptide are monitored, soluble and/or membrane bound forms of the polypeptide may be

utilized. Where membrane-bound forms of the polypeptide are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton[®] X-100, Triton[®] X-114, Thesit[®], Isotridecypoly(ethylene glycol ether)_n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

[0144] An interaction between a test molecule and target molecule also can be detected by monitoring fluorescence energy transfer (FET) (*see, e.g.,* Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos *et al.* U.S. Patent No. 4,868,103). A fluorophore label on a first, “donor” molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, “acceptor” molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the “donor” polypeptide molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the “acceptor” molecule label may be differentiated from that of the “donor”. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the “acceptor” molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (*e.g.,* using a fluorimeter).

[0145] In another embodiment, determining the presence or absence of an interaction between a test molecule and a target molecule can be effected by monitoring surface plasmon resonance (*see, e.g.,* Sjolander & Urbanicz, *Anal. Chem.* 63: 2338-2345 (1991) and Szabo *et al.*, *Curr. Opin. Struct. Biol.* 5: 699-705 (1995)). “Surface plasmon resonance” or “biomolecular interaction analysis (BIA)” can be utilized to detect biospecific interactions in real time, without labeling any of the interactants (*e.g.,* BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[0146] In another embodiment, the target molecule or test molecules are anchored to a solid phase, facilitating the detection of target molecule/test molecule complexes and separation of the complexes from free, uncomplexed molecules. The target molecule or test molecule is immobilized to the solid support. In an embodiment, the target molecule is anchored to a solid surface, and the test molecule, which is not anchored, can be labeled, either directly or indirectly, with detectable labels discussed herein.

[0147] It may be desirable to immobilize a target molecule, an anti-target molecule antibody, and/or test molecules to facilitate separation of target molecule/test molecule complexes from uncomplexed forms, as well as to accommodate automation of the assay. The attachment between a test molecule and/or target molecule and the solid support may be covalent or non-covalent (*see, e.g.*, U.S. Patent No. 6,022,688 for non-covalent attachments). The solid support may be one or more surfaces of the system, such as one or more surfaces in each well of a microtiter plate, a surface of a silicon wafer, a surface of a bead (*see, e.g.*, Lam, *Nature* 354: 82-84 (1991)) that is optionally linked to another solid support, or a channel in a microfluidic device, for example. Types of solid supports, linker molecules for covalent and non-covalent attachments to solid supports, and methods for immobilizing nucleic acids and other molecules to solid supports are well known (*see, e.g.*, U.S. Patent Nos. 6,261,776; 5,900,481; 6,133,436; and 6,022,688; and WIPO publication WO 01/18234).

[0148] In an embodiment, target molecule may be immobilized to surfaces via biotin and streptavidin. For example, biotinylated target polypeptide can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). In another embodiment, a target polypeptide can be prepared as a fusion polypeptide. For example, glutathione-S-transferase/target polypeptide fusion can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivitized microtiter plates, which are then combined with a test molecule under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, or the matrix is immobilized in the case of beads, and complex formation is determined directly or indirectly as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of target molecule binding or activity is determined using standard techniques.

[0149] In an embodiment, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) under conditions such that a significant percentage of complexes formed will remain immobilized to the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of manners. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface, *e.g.*, by adding a labeled antibody specific for the immobilized component, where the antibody, in turn, can be directly labeled or indirectly labeled with, *e.g.*, a labeled anti-Ig antibody.

[0150] In another embodiment, an assay is performed utilizing antibodies that specifically bind target molecule or test molecule but do not interfere with binding of the target molecule to the test

molecule. Such antibodies can be derivitized to a solid support, and unbound target molecule may be immobilized by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

[0151] Cell free assays also can be conducted in a liquid phase. In such an assay, reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (*see, e.g.,* Rivas, G., and Minton, *Trends Biochem Sci Aug;18(8): 284-7 (1993)*); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (*see, e.g.,* Ausubel *et al., eds. Current Protocols in Molecular Biology*, J. Wiley: New York (1999)); and immunoprecipitation (*see, e.g.,* Ausubel *et al., eds., supra*). Media and chromatographic techniques are known to one skilled in the art (*see, e.g.,* Heegaard, *J Mol. Recognit. Winter; 11(1-6): 141-8 (1998)*; Hage & Tweed, *J. Chromatogr. B Biomed. Sci. Appl. Oct 10; 699 (1-2): 499-525 (1997)*). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[0152] In another embodiment, modulators of target molecule expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of target mRNA or target polypeptide is evaluated relative to the level of expression of target mRNA or target polypeptide in the absence of the candidate compound. When expression of target mRNA or target polypeptide is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as an agonist of target mRNA or target polypeptide expression. Alternatively, when expression of target mRNA or target polypeptide is less (*e.g.,* less with statistical significance) in the presence of the candidate compound than in its absence, the candidate compound is identified as an antagonist or inhibitor of target mRNA or target polypeptide expression. The level of target mRNA or target polypeptide expression can be determined by methods described herein.

[0153] In another embodiment, binding partners that interact with a target molecule are detected. The target molecules can interact with one or more cellular or extracellular macromolecules, such as polypeptides *in vivo*, and these interacting molecules are referred to herein as "binding partners." Binding partners can agonize or antagonize target molecule biological activity. Also, test molecules that agonize or antagonize interactions between target molecules and binding partners can be useful as therapeutic molecules as they can up-regulate or down-regulated target molecule activity *in vivo* and thereby treat osteoarthritis.

[0154] Binding partners of target molecules can be identified by methods known in the art. For example, binding partners may be identified by lysing cells and analyzing cell lysates by electrophoretic techniques. Alternatively, a two-hybrid assay or three-hybrid assay can be utilized (*see, e.g.,* U.S. Patent

No. 5,283,317; Zervos *et al.*, *Cell* 72:223-232 (1993); Madura *et al.*, *J. Biol. Chem.* 268: 12046-12054 (1993); Bartel *et al.*, *Biotechniques* 14: 920-924 (1993); Iwabuchi *et al.*, *Oncogene* 8: 1693-1696 (1993); and Brent WO94/10300). A two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. The assay often utilizes two different DNA constructs. In one construct, a *TM7SF3* nucleic acid (sometimes referred to as the “bait”) is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In another construct, a DNA sequence from a library of DNA sequences that encodes a potential binding partner (sometimes referred to as the “prey”) is fused to a gene that encodes an activation domain of the known transcription factor. Sometimes, a *TM7SF3* nucleic acid can be fused to the activation domain. If the “bait” and the “prey” molecules interact *in vivo*, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to identify the potential binding partner.

[0155] In an embodiment for identifying test molecules that antagonize or agonize complex formation between target molecules and binding partners, a reaction mixture containing the target molecule and the binding partner is prepared, under conditions and for a time sufficient to allow complex formation. The reaction mixture often is provided in the presence or absence of the test molecule. The test molecule can be included initially in the reaction mixture, or can be added at a time subsequent to the addition of the target molecule and its binding partner. Control reaction mixtures are incubated without the test molecule or with a placebo. Formation of any complexes between the target molecule and the binding partner then is detected. Decreased formation of a complex in the reaction mixture containing test molecule as compared to in a control reaction mixture indicates that the molecule antagonizes target molecule/binding partner complex formation. Alternatively, increased formation of a complex in the reaction mixture containing test molecule as compared to in a control reaction mixture indicates that the molecule agonizes target molecule/binding partner complex formation. In another embodiment, complex formation of target molecule/binding partner can be compared to complex formation of mutant target molecule/binding partner (*e.g.*, amino acid modifications in a target polypeptide). Such a comparison can be important in those cases where it is desirable to identify test molecules that modulate interactions of mutant but not non-mutated target gene products.

[0156] The assays can be conducted in a heterogeneous or homogeneous format. In heterogeneous assays, target molecule and/or the binding partner are immobilized to a solid phase, and complexes are detected on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the molecules being tested. For example, test compounds that agonize target

molecule/binding partner interactions can be identified by conducting the reaction in the presence of the test molecule in a competition format. Alternatively, test molecules that agonize preformed complexes, *e.g.*, molecules with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed.

[0157] In a heterogeneous assay embodiment, the target molecule or the binding partner is anchored onto a solid surface (*e.g.*, a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored molecule can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the molecule to be anchored can be used to anchor the molecule to the solid surface. The partner of the immobilized species is exposed to the coated surface with or without the test molecule. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) such that a significant portion of any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface is indicative of complex. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored to the surface; *e.g.*, by using a labeled antibody specific for the initially non-immobilized species. Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

[0158] In another embodiment, the reaction can be conducted in a liquid phase in the presence or absence of test molecule, where the reaction products are separated from unreacted components, and the complexes are detected (*e.g.*, using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes). Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

[0159] In an alternate embodiment, a homogeneous assay can be utilized. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared. One or both of the target molecule or binding partner is labeled, and the signal generated by the label(s) is quenched upon complex formation (*e.g.*, U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). Addition of a test molecule that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target molecule/binding partner complexes can be identified.

[0160] Candidate therapeutics for treating osteoarthritis are identified from a group of test molecules that interact with a target molecule. Test molecules are normally ranked according to the degree with which they modulate (*e.g.*, agonize or antagonize) a function associated with the target molecule (*e.g.*, DNA replication and/or processing, RNA transcription and/or processing, polypeptide production and/or processing, and/or biological function/activity), and then top ranking modulators are selected. Also,

pharmacogenomic information described herein can determine the rank of a modulator. The top 10% of ranked test molecules often are selected for further testing as candidate therapeutics, and sometimes the top 15%, 20%, or 25% of ranked test molecules are selected for further testing as candidate therapeutics. Candidate therapeutics typically are formulated for administration to a subject.

Therapeutic Formulations

[0161] Formulations and pharmaceutical compositions typically include in combination with a pharmaceutically acceptable carrier one or more target molecule modulators. The modulator often is a test molecule identified as having an interaction with a target molecule by a screening method described above. The modulator may be a compound, an antisense nucleic acid, a ribozyme, an antibody, or a binding partner. Also, formulations may comprise a target polypeptide or fragment thereof in combination with a pharmaceutically acceptable carrier.

[0162] As used herein, the term “pharmaceutically acceptable carrier” includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions. Pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

[0163] A pharmaceutical composition typically is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerin, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[0164] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, *e.g.*, gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a

disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[0165] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0166] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0167] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

[0168] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal

administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art. Molecules can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[0169] In one embodiment, active molecules are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. Materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[0170] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[0171] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Molecules which exhibit high therapeutic indices are preferred. While molecules that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0172] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such molecules lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any molecules used in the methods described herein, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be

used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[0173] As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, sometimes about 0.01 to 25 mg/kg body weight, often about 0.1 to 20 mg/kg body weight, and more often about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, sometimes between 2 to 8 weeks, often between about 3 to 7 weeks, and more often for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

[0174] With regard to polypeptide formulations, featured herein is a method for treating osteoarthritis in a subject, which comprises contacting one or more cells in the subject with a first polypeptide, where the subject comprises a second polypeptide having one or more polymorphic variations associated with cancer, and where the first polypeptide comprises fewer polymorphic variations associated with cancer than the second polypeptide. The first and second polypeptides are encoded by a nucleic acid which comprises a nucleotide sequence in SEQ ID NO: 1-2; a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence referenced in SEQ ID NO: 1-2; a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2 and a nucleotide sequence 90% or more identical to a nucleotide sequence in SEQ ID NO: 1-2. The subject often is a human.

[0175] For antibodies, a dosage of 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg) is often utilized. If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is often appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (*e.g.*, into the brain). A method for lipidation of antibodies is described by Cruikshank *et al.*, *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193 (1997).

[0176] Antibody conjugates can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for

example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[0177] For compounds, exemplary doses include milligram or microgram amounts of the compound per kilogram of subject or sample weight, for example, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid described herein, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[0178] With regard to nucleic acid formulations, gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent 5,328,470) or by stereotactic injection (*see e.g.*, Chen *et al.*, (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). Pharmaceutical preparations of gene therapy vectors can include a gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells (*e.g.*, retroviral vectors) the pharmaceutical preparation can include one or more cells which produce the gene delivery system. Examples of gene delivery vectors are described herein.

Therapeutic Methods

[0179] A therapeutic formulation described above can be administered to a subject in need of a therapeutic for inducing a desired biological response.. Therapeutic formulations can be administered by any of the paths described herein. With regard to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from pharmacogenomic analyses described herein.

[0180] As used herein, the term “treatment” is defined as the application or administration of a therapeutic formulation to a subject, or application or administration of a therapeutic agent to an isolated tissue or cell line from a subject with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect osteoarthritis, symptoms of osteoarthritis or a predisposition towards osteoarthritis. A therapeutic formulation includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides. Administration of a therapeutic formulation can occur prior to the manifestation of symptoms characteristic of osteoarthritis, such that osteoarthritis is prevented or delayed in its progression. The appropriate therapeutic composition can be determined based on screening assays described herein.

[0181] As discussed, successful treatment of osteoarthritis can be brought about by techniques that serve to agonize target molecule expression or function, or alternatively, antagonize target molecule expression or function. These techniques include administration of modulators that include, but are not limited to, small organic or inorganic molecules; antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and Fab, F(ab')₂ and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof); and peptides, phosphopeptides, or polypeptides.

[0182] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above. It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular polypeptide, it can be preferable to co-administer normal target gene polypeptide into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

[0183] Another method by which nucleic acid molecules may be utilized in treating or preventing osteoarthritis is use of aptamer molecules specific for target molecules. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to ligands (*see, e.g., Osborne, et al., Curr. Opin. Chem. Biol.* 1(1): 5-9 (1997); and Patel, D. J., *Curr. Opin. Chem. Biol. Jun; 1(1): 32-46 (1997)*).

[0184] Yet another method of utilizing nucleic acid molecules for osteoarthritis treatment is gene therapy, which can also be referred to as allele therapy. Provided herein is a gene therapy method for

treating osteoarthritis in a subject, which comprises contacting one or more cells in the subject or from the subject with a nucleic acid having a first nucleotide sequence (e.g., the first nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1-2). Genomic DNA in the subject comprises a second nucleotide sequence having one or more polymorphic variations associated with osteoarthritis (e.g., the second nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1). The first and second nucleotide sequences typically are substantially identical to one another, and the first nucleotide sequence comprises fewer polymorphic variations associated with osteoarthritis than the second nucleotide sequence. The first nucleotide sequence may comprise a gene sequence that encodes a full-length polypeptide or a fragment thereof. The subject is often a human. Allele therapy methods often are utilized in conjunction with a method of first determining whether a subject has genomic DNA that includes polymorphic variants associated with osteoarthritis.

[0185] In another allele therapy embodiment, provided herein is a method which comprises contacting one or more cells in the subject or from the subject with a polypeptide encoded by a nucleic acid having a first nucleotide sequence (e.g., the first nucleotide sequence is identical to or substantially identical to the nucleotide sequence of SEQ ID NO: 1-2). Genomic DNA in the subject comprises a second nucleotide sequence having one or more polymorphic variations associated with osteoarthritis (e.g., the second nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1). The first and second nucleotide sequences typically are substantially identical to one another, and the first nucleotide sequence comprises fewer polymorphic variations associated with osteoarthritis than the second nucleotide sequence. The first nucleotide sequence may comprise a gene sequence that encodes a full-length polypeptide or a fragment thereof. The subject is often a human.

[0186] For antibody-based therapies, antibodies can be generated that are both specific for target molecules and that reduce target molecule activity. Such antibodies may be administered in instances where antagonizing a target molecule function is appropriate for the treatment of osteoarthritis.

[0187] In circumstances where stimulating antibody production in an animal or a human subject by injection with a target molecule is harmful to the subject, it is possible to generate an immune response against the target molecule by use of anti-idiotypic antibodies (*see, e.g., Herlyn, Ann. Med.; 31(1): 66-78 (1999); and Bhattacharya-Chatterjee & Foon, Cancer Treat. Res.; 94: 51-68 (1998)*). Introducing an anti-idiotypic antibody to a mammal or human subject often stimulates production of anti-anti-idiotypic antibodies, which typically are specific to the target molecule. Vaccines directed to osteoarthritis also may be generated in this fashion.

[0188] In instances where the target molecule is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody

are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (*see, e.g., Marasco et al., Proc. Natl. Acad. Sci. USA 90: 7889-7893 (1993)*).

[0189] Modulators can be administered to a patient at therapeutically effective doses to treat osteoarthritis. A therapeutically effective dose refers to an amount of the modulator sufficient to result in amelioration of symptoms of osteoarthritis. Toxicity and therapeutic efficacy of modulators can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.,* for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Modulators that exhibit large therapeutic indices are preferred. While modulators that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such molecules to the site of affected tissue in order to minimize potential damage to uninfected cells, thereby reducing side effects.

[0190] Data obtained from cell culture assays and animal studies can be used in formulating a range of dosages for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the methods described herein, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (*i.e.,* the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[0191] Another example of effective dose determination for an individual is the ability to directly assay levels of “free” and “bound” compound in the serum of the test subject. Such assays may utilize antibody mimics and/or “biosensors” that have been created through molecular imprinting techniques. Molecules that modulate target molecule activity are used as a template, or “imprinting molecule”, to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated “negative image” of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell *et al., Current Opinion in Biotechnology 7: 89-94 (1996)* and in Shea, *Trends in Polymer Science 2: 166-173 (1994)*. Such

“imprinted” affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, *et al.*, *Nature* 361: 645-647 (1993). Through the use of isotope-labeling, the “free” concentration of compound which modulates target molecule expression or activity readily can be monitored and used in calculations of IC_{50} . Such “imprinted” affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes readily can be assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC_{50} . An example of such a “biosensor” is discussed in Kriz *et al.*, *Analytical Chemistry* 67: 2142-2144 (1995).

[0192] The examples set forth below are intended to illustrate but not limit the invention.

Examples

[0193] In the following studies a group of subjects was selected according to specific parameters relating to osteoarthritis (OA). Nucleic acid samples obtained from individuals in the study group were subjected to genetic analysis, which identified associations between osteoarthritis and a polymorphism in the *TM7SF3* gene on chromosome twelve. The polymorphism was genotyped again in two replication cohorts consisting of individuals selected for OA. In addition, SNPs proximal to the incident polymorphism were identified and allelotyped in OA case and control pools. Methods are described for producing *TM7SF3* polypeptide and *TM7SF3* polypeptide variants *in vitro* or *in vivo*, *TM7SF3* nucleic acids or polypeptides and variants thereof are utilized for screening test molecules for those that interact with *TM7SF3* molecules. Test molecules identified as interactors with *TM7SF3* molecules and *TM7SF3* variants are further screened *in vivo* to determine whether they treat osteoarthritis.

Example 1

Samples and Pooling Strategies

Sample Selection

[0194] Blood samples were collected from individuals diagnosed with knee osteoarthritis, which were referred to as case samples. Also, blood samples were collected from individuals not diagnosed with knee osteoarthritis as gender and age-matched controls. A database was created that listed all phenotypic trait information gathered from individuals for each case and control sample. Genomic DNA was extracted from each of the blood samples for genetic analyses.

DNA Extraction from Blood Samples

[0195] Six to ten milliliters of whole blood was transferred to a 50 ml tube containing 27 ml of red cell lysis solution (RCL). The tube was inverted until the contents were mixed. Each tube was incubated for 10 minutes at room temperature and inverted once during the incubation. The tubes were then centrifuged for 20 minutes at 3000 x g and the supernatant was carefully poured off. 100-200 µl of residual liquid was left in the tube and was pipetted repeatedly to resuspend the pellet in the residual supernatant. White cell lysis solution (WCL) was added to the tube and pipetted repeatedly until completely mixed. While no incubation was normally required, the solution was incubated at 37°C or room temperature if cell clumps were visible after mixing until the solution was homogeneous. 2 ml of protein precipitation was added to the cell lysate. The mixtures were vortexed vigorously at high speed for 20 sec to mix the protein precipitation solution uniformly with the cell lysate, and then centrifuged for 10 minutes at 3000 x g. The supernatant containing the DNA was then poured into a clean 15 ml tube, which contained 7 ml of 100% isopropanol. The samples were mixed by inverting the tubes gently until white threads of DNA were visible. Samples were centrifuged for 3 minutes at 2000 x g and the DNA was visible as a small white pellet. The supernatant was decanted and 5 ml of 70% ethanol was added to each tube. Each tube was inverted several times to wash the DNA pellet, and then centrifuged for 1 minute at 2000 x g. The ethanol was decanted and each tube was drained on clean absorbent paper. The DNA was dried in the tube by inversion for 10 minutes, and then 1000 µl of 1X TE was added. The size of each sample was estimated, and less TE buffer was added during the following DNA hydration step if the sample was smaller. The DNA was allowed to rehydrate overnight at room temperature, and DNA samples were stored at 2-8°C.

[0196] DNA was quantified by placing samples on a hematology mixer for at least 1 hour. DNA was serially diluted (typically 1:80, 1:160, 1:320, and 1:640 dilutions) so that it would be within the measurable range of standards. 125 µl of diluted DNA was transferred to a clear U-bottom microtitre plate, and 125 µl of 1X TE buffer was transferred into each well using a multichannel pipette. The DNA and 1X TE were mixed by repeated pipetting at least 15 times, and then the plates were sealed. 50 µl of diluted DNA was added to wells A5-H12 of a black flat bottom microtitre plate. Standards were inverted six times to mix them, and then 50 µl of 1X TE buffer was pipetted into well A1, 1000 ng/ml of standard was pipetted into well A2, 500 ng/ml of standard was pipetted into well A3, and 250 ng/ml of standard was pipetted into well A4. PicoGreen (Molecular Probes, Eugene, Oregon) was thawed and freshly diluted 1:200 according to the number of plates that were being measured. PicoGreen was vortexed and then 50µl was pipetted into all wells of the black plate with the diluted DNA. DNA and PicoGreen were mixed by pipetting repeatedly at least 10 times with the multichannel pipette. The plate was placed into a Fluoroskan Ascent Machine (microplate fluorometer produced by Labsystems) and the samples were

allowed to incubate for 3 minutes before the machine was run using filter pairs 485 nm excitation and 538 nm emission wavelengths. Samples having measured DNA concentrations of greater than 450 ng/μl were re-measured for conformation. Samples having measured DNA concentrations of 20 ng/μl or less were re-measured for confirmation.

Pooling Strategies – Discovery Cohort

[0197] Samples were derived from the Nottingham knee OA family study (UK) where index cases were identified through a knee replacement registry. Siblings were approached and assessed with knee x-rays and assigned status as affected or unaffected. In all 1,157 individuals were available. In order to create same-sex pools of appropriate sizes, 335 unrelated female individuals with OA from the Nottingham OA sample were selected for the case pool. The control pool was made up of unrelated female individuals from the St. Thomas twin study (England) with normal knee x-rays and without other indications of OA, regardless of anatomical location, as well as lacking family history of OA. The St. Thomas twin study consists of Caucasian, female participants from the St. Thomas' Hospital, London, adult-twin registry, which is a voluntary registry of >4,000 twin pairs ranging from 18 to 76 years of age. The female case samples and female control samples are described further in Table 1 below.

[0198] A select set of samples from each group were utilized to generate pools, and one pool was created for each group. Each individual sample in a pool was represented by an equal amount of genomic DNA. For example, where 25 ng of genomic DNA was utilized in each PCR reaction and there were 200 individuals in each pool, each individual would provide 125 pg of genomic DNA. Inclusion or exclusion of samples for a pool was based upon the following criteria: the sample was derived from an individual characterized as Caucasian; the sample was derived from an individual of British paternal and maternal descent; case samples were derived from individuals diagnosed with specific knee osteoarthritis (OA) and were recruited from an OA knee replacement clinic. Control samples were derived from individuals free of OA, family history of OA, and rheumatoid arthritis. Also, sufficient genomic DNA was extracted from each blood sample for all allelotyping and genotyping reactions performed during the study. Phenotype information from each individual was collected and included age of the individual, gender, family history of OA, general medical information (e.g., height, weight, thyroid disease, diabetes, psoriasis, hysterectomy), joint history (previous and current symptoms, joint-related operations, age at onset of symptoms, date of primary diagnosis, age of individual as of primary diagnosis and order of involvement), and knee-related findings (crepitus, restricted passive movement, bony swelling/deformity). Additional knee information included knee history, current symptoms, any major knee injury, menisectomy, knee replacement surgery, age of surgery, and treatment history (including

hormone replace therapy (HRT)). Samples that met these criteria were added to appropriate pools based on disease status.

[0199] The selection process yielded the pools set forth in Table 1, which were used in the studies that follow:

TABLE 1

	Female case	Female control
Pool size (Number)	335	335
Pool Criteria (ex: case/control)	control	case
Mean Age (ex: years)	57.21	69.95

Example 2

Association of Polymorphic Variants with Osteoarthritis

[0200] A whole-genome screen was performed to identify particular SNPs associated with occurrence of osteoarthritis. As described in Example 1, two sets of samples were utilized, which included samples from female individuals having knee osteoarthritis (osteoarthritis cases), and samples from female individuals not having knee osteoarthritis (female controls). The initial screen of each pool was performed in an allelotyping study, in which certain samples in each group were pooled. By pooling DNA from each group, an allele frequency for each SNP in each group was calculated. These allele frequencies were then compared to one another. Particular SNPs were considered as being associated with osteoarthritis when allele frequency differences calculated between case and control pools were statistically significant. SNP disease association results obtained from the allelotyping study were then validated by genotyping each associated SNP across all samples from each pool. The results of the genotyping then were analyzed, allele frequencies for each group were calculated from the individual genotyping results, and a p-value was calculated to determine whether the case and control groups had statistically significant differences in allele frequencies for a particular SNP. When the genotyping results agreed with the original allelotyping results, the SNP disease association was considered validated at the genetic level.

SNP Panel Used for Genetic Analyses

[0201] A whole-genome SNP screen began with an initial screen of approximately 25,000 SNPs over each set of disease and control samples using a pooling approach. The pools studied in the screen are described in Example 1. The SNPs analyzed in this study were part of a set of 25,488 SNPs

confirmed as being statistically polymorphic as each is characterized as having a minor allele frequency of greater than 10%. The SNPs in the set reside in genes or in close proximity to genes, and many reside in gene exons. Specifically, SNPs in the set are located in exons, introns, and within 5,000 base-pairs upstream of a transcription start site of a gene. In addition, SNPs were selected according to the following criteria: they are located in ESTs; they are located in Locuslink or Ensembl genes; and they are located in Genomatix promoter predictions. SNPs in the set were also selected on the basis of even spacing across the genome, as depicted in Table 2.

[0202] A case-control study design using a whole genome association strategy involving approximately 28,000 single nucleotide polymorphisms (SNPs) was employed. Approximately 25,000 SNPs were evenly spaced in gene-based regions of the human genome with a median inter-marker distance of about 40,000 base pairs. Additionally, approximately 3,000 SNPs causing amino acid substitutions in genes described in the literature as candidates for various diseases were used. The case-control study samples were of female Caucasian origin (British paternal and maternal descent) 670 individuals were equally distributed in two groups: female controls and female cases. The whole genome association approach was first conducted on 2 DNA pools representing the 2 groups. Significant markers were confirmed by individual genotyping.

TABLE 2

<u>General Statistics</u>		<u>Spacing Statistics</u>	
Total # of SNPs	25,488	Median	37,058 bp
# of Exonic SNPs	>4,335 (17%)	Minimum*	1,000 bp
# SNPs with refSNP ID	20,776 (81%)	Maximum*	3,000,000 bp
Gene Coverage	>10,000	Mean	122,412 bp
Chromosome Coverage	All	Std Deviation	373,325 bp
		<i>*Excludes outliers</i>	

Allelotyping and Genotyping Results

[0203] The genetic studies summarized above and described in more detail below identified an allelic variant in the TM7SF3 gene that is associated with osteoarthritis.

Assay for Verifying, Allelotyping, and Genotyping SNPs

[0204] A MassARRAY™ system (Sequenom, Inc.) was utilized to perform SNP genotyping in a high-throughput fashion. This genotyping platform was complemented by a homogeneous, single-tube assay method (hME™ or homogeneous MassEXTEND™ (Sequenom, Inc.)) in which two genotyping primers anneal to and amplify a genomic target surrounding a polymorphic site of interest. A third

primer (the MassEXTEND™ primer), which is complementary to the amplified target up to but not including the polymorphism, was then enzymatically extended one or a few bases through the polymorphic site and then terminated.

[0205] For each polymorphism, SpectroDESIGNER™ software (Sequenom, Inc.) was used to generate a set of PCR primers and a MassEXTEND™ primer which were used to genotype the polymorphism. Other primer design software could be used or one of ordinary skill in the art could manually design primers based on his or her knowledge of the relevant factors and considerations in designing such primers. Table 3 shows PCR primers and Table 4 shows extension primers used for analyzing polymorphisms. The initial PCR amplification reaction was performed in a 5 µl total volume containing 1X PCR buffer with 1.5 mM MgCl₂ (Qiagen), 200 µM each of dATP, dGTP, dCTP, dTTP (Gibco-BRL), 2.5 ng of genomic DNA, 0.1 units of HotStar DNA polymerase (Qiagen), and 200 nM each of forward and reverse PCR primers specific for the polymorphic region of interest.

TABLE 3: PCR Primers

SNP Reference	Forward PCR primer	Reverse PCR primer
rs1484086	ACGTTGGATGTGTCACTCTTCGGAAGTCTC	ACGTTGGATGCATGTACAGGGCATTACAG

[0206] Samples were incubated at 95°C for 15 minutes, followed by 45 cycles of 95°C for 20 seconds, 56°C for 30 seconds, and 72°C for 1 minute, finishing with a 3 minute final extension at 72°C. Following amplification, shrimp alkaline phosphatase (SAP) (0.3 units in a 2 µl volume) (Amersham Pharmacia) was added to each reaction (total reaction volume was 7 µl) to remove any residual dNTPs that were not consumed in the PCR step. Samples were incubated for 20 minutes at 37°C, followed by 5 minutes at 85°C to denature the SAP.

[0207] Once the SAP reaction was complete, a primer extension reaction was initiated by adding a polymorphism-specific MassEXTEND™ primer cocktail to each sample. Each MassEXTEND™ cocktail included a specific combination of dideoxynucleotides (ddNTPs) and deoxynucleotides (dNTPs) used to distinguish polymorphic alleles from one another. Methods for verifying, allelotyping and genotyping SNPs are disclosed, for example, in U.S. Pat. No. 6,258,538, the content of which is hereby incorporated by reference. In Table 4, ddNTPs are shown and the fourth nucleotide not shown is the dNTP.

TABLE 4: Extension Primers

SNP Reference	Extend Probe	Termination Mix
rs1484086	CTCTTCGGAAGTCTCTTTCTCA	ACT

[0208] The MassEXTEND™ reaction was performed in a total volume of 9 µl, with the addition of 1X ThermoSequenase buffer, 0.576 units of ThermoSequenase (Amersham Pharmacia), 600 nM MassEXTEND™ primer, 2 mM of ddATP and/or ddCTP and/or ddGTP and/or ddTTP, and 2 mM of dATP or dCTP or dGTP or dTTP. The deoxy nucleotide (dNTP) used in the assay normally was complementary to the nucleotide at the polymorphic site in the amplicon. Samples were incubated at 94°C for 2 minutes, followed by 55 cycles of 5 seconds at 94°C, 5 seconds at 52°C, and 5 seconds at 72°C.

[0209] Following incubation, samples were desalted by adding 16 µl of water (total reaction volume was 25 µl), 3 mg of SpectroCLEAN™ sample cleaning beads (Sequenom, Inc.) and allowed to incubate for 3 minutes with rotation. Samples were then robotically dispensed using a piezoelectric dispensing device (SpectroJET™ (Sequenom, Inc.)) onto either 96-spot or 384-spot silicon chips containing a matrix that crystallized each sample (SpectroCHIP™ (Sequenom, Inc.)). Subsequently, MALDI-TOF mass spectrometry (Biflex and Autoflex MALDI-TOF mass spectrometers (Bruker Daltonics) can be used) and SpectroTYPER RT™ software (Sequenom, Inc.) were used to analyze and interpret the SNP genotype for each sample.

Genetic Analysis

[0210] Minor allelic frequencies for the polymorphisms set forth in Table A were verified as being 10% or greater using the extension assay described above in a group of samples isolated from 92 individuals originating from the state of Utah in the United States, Venezuela and France (Coriell cell repositories).

[0211] Genotyping results are shown for female pools in Table 5. In Table 5, “AF” refers to allelic frequency; and “F case” and “F control” refer to female case and female control groups, respectively.

TABLE 5: Genotyping Results

SNP Reference	AF F case	AF F control	p-value
rs1484086	T = 0.821 C = 0.179	T = 0.753 C = 0.247	0.0027

[0212] The single marker alleles set forth in Table A were considered validated, since the genotyping data agreed with the allelotyping data and each SNP significantly associated with osteoarthritis. Particularly significant associations with osteoarthritis are indicated by a calculated p-value of less than 0.05 for genotype results.

Example 3

Association of Polymorphic Variants with Osteoarthritis in Replication Cohorts

[0213] The single marker polymorphism set forth in Table A was genotyped again in two replication cohorts consisting of individuals selected for OA.

Sample Selection and Pooling Strategies – Replication Sample 1

[0214] A second case control sample (replication sample #1) was created by using 100 Caucasian female cases from Chingford, UK, and 148 unrelated female cases from the St. Thomas twin study. Cases were defined as having Kellgren-Lawrence (KL) scores of at least 2 in at least one knee x-ray. In addition, 199 male knee replacement cases from Nottingham were included. (For a cohort description, see the Nottingham description provided in Example 1). The control pool was made up of unrelated female individuals from the St. Thomas twin study (England) with normal knee x-rays and without other indications of OA, regardless of anatomical location, as well as lacking family history of OA. The St. Thomas twin study consists of Caucasian, female participants from the St. Thomas' Hospital, London, adult-twin registry, which is a voluntary registry of >4,000 twin pairs ranging from 18 to 76 years of age. The replication sample 1 cohort was used to replicate the initial results. Table 6 below summarizes the selected phenotype data collected from the case and control individuals.

TABLE 6

Phenotype	Female cases (n=248): median (range)/ (n,%)	Male cases (n=199): median (range)/ (n,%)	Female controls (n=313): mean (range)/ (n,%)
Age	59 (39- 73)	66 (45- 73)	55 (50- 72)
Height (cm)	162 (141- 178)	175 (152- 198)	162 (141- 176)
Weight (kg)	68 (51- 123)	86 (62- 127)	64 (40- 111)
Body mass index (kg/m ²)	26 (18- 44)	29 (21- 41)	24 (18- 46)
Kellgren- Lawrence* left knee	0 (63, 26%), 1 (20, 8%), 2 (105, 43%), 3 (58, 23%), 4 (1, 0%)	NA	NA
Kellgren- Lawrence* right knee	0 (43, 7%), 1 (18, 7%), 2 (127, 52%), 3 (57, 23%), 4 (1, 0%)	NA	NA
KL* >2 both knees	No (145, 59%), Yes (101, 41%)	NA	NA
KL* >2 either knee	No (0, 0%), Yes (248, 100%)	NA	NA

* 0: normal, 1: doubtful, 2: definite osteophyte (bony protuberance), 3: joint space narrowing (with or without osteophyte), 4: joint deformity

Sample Selection and Pooling Strategies – Replication Sample 2

[0215] A third case control sample (replication sample #2) was created by using individuals with symptoms of OA. These individuals were recruited and examined by rheumatologists. Affected joints were x-rayed and a diagnosis of definite or probable OA was made according to American College of Rheumatology criteria by a single rheumatologist. Controls were recruited from volunteers without any symptoms from the musculoskeletal system based on a normal joint exam performed by a rheumatologist. Only cases with a diagnosis of definite OA were included in the study. Only individuals of Caucasian origin were included. The cases consisted of 228 individuals with definite knee OA, 106 individuals with definite hip OA, and 74 individuals with hip OA.

TABLE 7

Phenotype	Case	Control
Age at Visit	62.7	52.5
Sex (Female/Male)	227/119	174/101
Knee OA Xray: No	35% (120)	80% (16)
Unknown	1% (4)	0% (0)
Yes	64% (221)	20% (4)
Hip OA Xray: No	63% (215)	80% (16)
Unknown	2% (7)	0% (0)
Yes	35% (121)	20% (4)

Assay for Verifying, Allelotyping, and Genotyping SNPs

[0216] Genotyping of the replication cohorts described in Tables 6 and 7 was performed using the same methods used for the original genotyping, as described herein. A MassARRAY™ system (Sequenom, Inc.) was utilized to perform SNP genotyping in a high-throughput fashion. This genotyping platform was complemented by a homogeneous, single-tube assay method (hME™ or homogeneous MassEXTEND™ (Sequenom, Inc.)) in which two genotyping primers anneal to and amplify a genomic target surrounding a polymorphic site of interest. A third primer (the MassEXTEND™ primer), which is complementary to the amplified target up to but not including the polymorphism, was then enzymatically extended one or a few bases through the polymorphic site and then terminated.

[0217] For each polymorphism, SpectroDESIGNER™ software (Sequenom, Inc.) was used to generate a set of PCR primers and a MassEXTEND™ primer which were used to genotype the

polymorphism. Other primer design software could be used or one of ordinary skill in the art could manually design primers based on his or her knowledge of the relevant factors and considerations in designing such primers. Table 3 shows PCR primers and Table 4 shows extension probes used for analyzing (*e.g.*, genotyping) polymorphisms in the replication cohorts. The initial PCR amplification reaction was performed in a 5 μ l total volume containing 1X PCR buffer with 1.5 mM MgCl₂ (Qiagen), 200 μ M each of dATP, dGTP, dCTP, dTTP (Gibco-BRL), 2.5 ng of genomic DNA, 0.1 units of HotStar DNA polymerase (Qiagen), and 200 nM each of forward and reverse PCR primers specific for the polymorphic region of interest.

[0218] Samples were incubated at 95°C for 15 minutes, followed by 45 cycles of 95°C for 20 seconds, 56°C for 30 seconds, and 72°C for 1 minute, finishing with a 3 minute final extension at 72°C. Following amplification, shrimp alkaline phosphatase (SAP) (0.3 units in a 2 μ l volume) (Amersham Pharmacia) was added to each reaction (total reaction volume was 7 μ l) to remove any residual dNTPs that were not consumed in the PCR step. Samples were incubated for 20 minutes at 37°C, followed by 5 minutes at 85°C to denature the SAP.

[0219] Once the SAP reaction was complete, a primer extension reaction was initiated by adding a polymorphism-specific MassEXTEND™ primer cocktail to each sample. Each MassEXTEND™ cocktail included a specific combination of dideoxynucleotides (ddNTPs) and deoxynucleotides (dNTPs) used to distinguish polymorphic alleles from one another. Methods for verifying, allelotyping and genotyping SNPs are disclosed, for example, in U.S. Pat. No. 6,258,538, the content of which is hereby incorporated by reference. In Table 7, ddNTPs are shown and the fourth nucleotide not shown is the dNTP.

[0220] The MassEXTEND™ reaction was performed in a total volume of 9 μ l, with the addition of 1X ThermoSequenase buffer, 0.576 units of ThermoSequenase (Amersham Pharmacia), 600 nM MassEXTEND™ primer, 2 mM of ddATP and/or ddCTP and/or ddGTP and/or ddTTP, and 2 mM of dATP or dCTP or dGTP or dTTP. The deoxy nucleotide (dNTP) used in the assay normally was complementary to the nucleotide at the polymorphic site in the amplicon. Samples were incubated at 94°C for 2 minutes, followed by 55 cycles of 5 seconds at 94°C, 5 seconds at 52°C, and 5 seconds at 72°C.

[0221] Following incubation, samples were desalted by adding 16 μ l of water (total reaction volume was 25 μ l), 3 mg of SpectroCLEAN™ sample cleaning beads (Sequenom, Inc.) and allowed to incubate for 3 minutes with rotation. Samples were then robotically dispensed using a piezoelectric dispensing device (SpectroJET™ (Sequenom, Inc.)) onto either 96-spot or 384-spot silicon chips containing a matrix that crystallized each sample (SpectroCHIP™ (Sequenom, Inc.)). Subsequently, MALDI-TOF mass

spectrometry (Biflex and Autoflex MALDI-TOF mass spectrometers (Bruker Daltonics) can be used) and SpectroTYPER RT™ software (Sequenom, Inc.) were used to analyze and interpret the SNP genotype for each sample.

Genetic Analysis

[0222] Genotyping results for replication cohorts #1 and #2 are provided in Tables 8 and 9, respectively.

TABLE 8

rsID	Replication #1 (Mixed Male/Female cases and Female controls)				Meta-analysis Disc. + Rep #1
	AF OA Con	AF OA Cas	Delta	P-value	P-value
rs1484086	0.23	0.20	0.03	0.287	0.0077

TABLE 9

rsID	Replication #2 (Newfoundland) (Male/Female cases and controls)				Meta-analysis Disc. + Rep #2
	AF OA Con	AF OA Cas	Delta	P-value	Not Done
rs1484086	0.23	0.21	0.026	0.280	

[0223] To combine the evidence for association from multiple sample collections, a meta-analysis procedure was employed. The allele frequencies were compared between cases and controls within the discovery sample, as well as within the replication cohort #1 using the DerSimian-Laird approach (DerSimonian, R. and N. Laird. 1986. Meta-analysis in clinical trials. Control Clin Trials 7: 177-188.)

[0224] The absence of a statistically significant association in one or more of the replication cohorts should not be interpreted as minimizing the value of the original finding. There are many reasons why a biologically derived association identified in a sample from one population would not replicate in a sample from another population. The most important reason is differences in population history. Due to bottlenecks and founder effects, there may be common disease predisposing alleles present in one population that are relatively rare in another, leading to a lack of association in the candidate region. Also, because common diseases such as arthritis-related disorders are the result of susceptibilities in many genes and many environmental risk factors, differences in population-specific genetic and environmental backgrounds could mask the effects of a biologically relevant allele. For these and other reasons, statistically strong results in the original, discovery sample that did not replicate in one or more of the replication samples may be further evaluated in additional replication cohorts and experimental systems.

Example 4

TM7SF3 Region Proximal SNPs

[0225] It has been discovered that SNP rs1484086 in *TM7SF3* is associated with occurrence of osteoarthritis in subjects. *TM7SF3* is an orphan receptor and is a member of the superfamily of 7-transmembrane domain proteins, one of the largest superfamilies of cell surface proteins. Members of this family include receptors for a variety of ligands, such as peptides, hormones, and ions, and for external sensory stimuli, such as odorants and light. Many 7-transmembrane molecules are able to recruit small G proteins, suggesting that they can transduce external signals to the cytoplasm.

[0226] Thirty-seven additional allelic variants proximal to rs1484086 were identified and subsequently allelotyped in osteoarthritis case and control sample sets as described in Examples 1 and 2. The polymorphic variants are set forth in Table 10. The chromosome positions provided in column four of Table 10 are based on Genome “Build 34” of NCBI’s GenBank.

TABLE 10

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs1058701	12	230	27004780	a/c
rs11613	12	231	27004781	c/g
rs900743	12	5330	27009880	g/t
rs2129091	12	6334	27010884	a/c
rs15556	12	11372	27015922	c/t
rs1053724	12	11456	27016006	g/t
rs9699	12	11501	27016051	a/g
rs4856	12	13393	27017943	c/t
rs3782314	12	16666	27021216	a/g
rs2087736	12	17596	27022146	c/t
rs2068372	12	19710	27024260	a/g
rs2068371	12	19800	27024350	a/g
rs1184529	12	20297	27024847	a/g
rs2101256	12	20967	27025517	g/t
rs1552257	12	32514	27037064	c/t
rs1565585	12	33159	27037709	a/g
rs4080800	12	37600	27042150	a/g
rs1388658	12	41259	27045809	a/g
rs3782316	12	41329	27045879	c/t
rs1484086	12	50060	27054610	c/t
rs1484087	12	53292	27057842	c/t
rs2291549	12	53393	27057943	a/g
rs1565584	12	56417	27060967	c/t
rs3071166	12	56435	27060985	-/tt
rs1907652	12	58847	27063397	a/t
rs3759119	12	59595	27064145	c/t
rs3759120	12	59661	27064211	g/t

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs3782317	12	60355	27064905	c/t
rs3825166	12	60407	27064957	a/c
rs1872191	12	62357	27066907	a/g
rs8628	12	68230	27072780	g/t
rs1872193	12	68516	27073066	a/g
rs1042833	12	69055	27073605	c/t
rs2476	12	72603	27077153	c/g
rs1976206	12	73928	27078478	a/g
rs1388659	12	85897	27090447	c/t
rs1184528	12	91554	27096104	a/g

Assay for Verifying and Allelotyping SNPs

[0227] The methods used to verify and allelotype the 37 proximal SNPs of Table 10 are the same methods described in Examples 1 and 2 herein. The primers and probes used in these assays are provided in Table 11 and Table 12, respectively.

TABLE 11

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs1058701	ACGTTGGATGCTTCTCTCCAGTCCATGTTG	ACGTTGGATGGTAACAAGTCCGAAGGATTG
rs11613	ACGTTGGATGTCCAAGTGCCCTTCTCTCCAG	ACGTTGGATGTCCCTTGATGCATCTCGACAC
rs900743	ACGTTGGATGAACAGATCCGGAACCTTTTT	ACGTTGGATGGGCTCATAGAACCCTTTTT
rs2129091	ACGTTGGATGTTAAATTTGGTATGGTCTCC	ACGTTGGATGGTTTTGCACTAAGAAGAGAC
rs15556	ACGTTGGATGTGTAGGGACAAAGTTATATG AA	ACGTTGGATGTGTCTTTACAGTCTTCACTGGC A
rs1053724	ACGTTGGATGCCATATAACTTTGTCCCTAC	ACGTTGGATGAAAATACAGTGCAGGTGACC
rs9699	ACGTTGGATGTTTCAGTGTTACTGAAGTTAAT T	ACGTTGGATGCAACATAGAAAATATAAAACAA TTT
rs4856	ACGTTGGATGTTGTAAACCAAAAGGTAATTC TCA	ACGTTGGATGGATGACCAAAAGGTTGGTGG
rs3782314	ACGTTGGATGAATGCTGCACAACCTTAGGCC	ACGTTGGATGAGCAGTCAGCCTATTCTTGC
rs2087736	ACGTTGGATGCAGTGAGACTCTGTCTCTAC	ACGTTGGATGTACATCAGCCTCCCAAGTAG
rs2068372	ACGTTGGATGCTAGAGAGGGTTGCTCTATG	ACGTTGGATGCCAGCTTCAGCTTTCTGTAG
rs2068371	ACGTTGGATGATTGACAAGCCAAGCCTGTC	ACGTTGGATGATAGAGCAACCCTCTCTAGG
rs1184529	ACGTTGGATGTTCCCTTCTCTGGTGGTTTGC	ACGTTGGATGGGTGTCAGTCTTGTAATCAG
rs2101256	ACGTTGGATGCTGGGTTTTTTTGGTGTGTTG	ACGTTGGATGCTTGTACTGTATCTCAATTTC
rs1552257	ACGTTGGATGCAATACCTTAGCTCACCTAC	ACGTTGGATGCATTAGGGAAAATCTGGATC
rs1565585	ACGTTGGATGGCTAGAAATATACCACAGGTG	ACGTTGGATGTCACTGAAAATGGATCAGCC
rs4080800	ACGTTGGATGGGGTTCAAGCAATTCTCCTG	ACGTTGGATGAAAATTAGCCAGGTGTGGTG
rs1388658	ACGTTGGATGGAAGAGAGAGGATGGGTG	ACGTTGGATGCAGGAGCTTAAGAACCACTG
rs3782316	ACGTTGGATGGTCTTGCTGTAATGTCATTAG	ACGTTGGATGCTTCTCTTTTTTAAGTAGATC
rs1484086	ACGTTGGATGTGTCACTCTTCGGAAGTCTC	ACGTTGGATGCATGTACAGGGCATTACAG
rs1484087	ACGTTGGATGGCATGGAATTTTACACCCC	ACGTTGGATGTTCCCTCTCAAACCTCTGTAG
rs2291549	ACGTTGGATGACTAAAGCGCCTCTTTGCTG	ACGTTGGATGCATTAAGCCGCTAGTTCCAC
rs1565584	ACGTTGGATGGAAGGCAAGTTATGATGC	ACGTTGGATGAGTATCCATTCTCTGAGTC

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs3071166	ACGTTGGATGGAAAAGGCAAGTTATGATGC	ACGTTGGATGAGTATCCATTCCTCTGAGTC
rs1907652	ACGTTGGATGCCATATTTCCAGTACCTAGC	ACGTTGGATGGGTTGATAAATATGTGCCAG
rs3759119	ACGTTGGATGCTAAACTGTTCTTAGTGCCG	ACGTTGGATGGGGAATTTCAAAGAGTGTGG
rs3759120	ACGTTGGATGCGGCACTAAGAACAGTTTAGA	ACGTTGGATGGGTTTTTATGACTGTAGCAAC
rs3782317	ACGTTGGATGGGAAGCTCTTGAAGCTGTAG	ACGTTGGATGGCCATTGAGAAATCCTGAGC
rs3825166	ACGTTGGATGACCAGACTCTCAACTTAGCC	ACGTTGGATGTGGCAGGTGGGTTATTCTTG
rs1872191	ACGTTGGATGTTTATGCCGAAGCCCTGTCT	ACGTTGGATGTAAAGCAAGGGAGGAAAGGG
rs8628	ACGTTGGATGCAAGCCTTACCAACAATTACA GAA	ACGTTGGATGGCTTATCAAGAGTGAAAATAGA AGA
rs1872193	ACGTTGGATGAATGCCACCTCTAAGAGGCA	ACGTTGGATGCTCAAGCCAAAGGAGAAGAC
rs1042833	ACGTTGGATGTCATCCAACCTTGGATCTC	ACGTTGGATGATACAGCCCCAGGAAATAGC
rs2476	ACGTTGGATGAGGAGCAGGTGACGTTAATG	ACGTTGGATGATTGACTAGCCACCAGGAAG
rs1976206	ACGTTGGATGAGCTGGGATTGGATTACAGG	ACGTTGGATGATGGAGAAACCCGCTCTCTA
rs1388659	ACGTTGGATGACCTCAGCCTTCCACATTTG	ACGTTGGATGAGCAGGAGGAAACTTTTGGG
rs1184528	ACGTTGGATGTGAGAATCGCTTGAACCCAG	ACGTTGGATGTTTGAATCAGAGTCTCCCTC

TABLE 12

dbSNP rs#	Extend Primer	Term Mix
rs1058701	CCAGTCCATGTTGAGGTGC	ACT
rs11613	CCAGTCCATGTTGAGGTG	ACT
rs900743	ATCCGGAACTTTTTTTTTGAAT	ACT
rs2129091	ATTTAATTTGAAAAGTGCTTACCC	ACT
rs15556	GTTTGTTATTTTCTATAC	ACG
rs1053724	CATCAAGTTTCAGTGTTA	CGT
rs9699	TACAGAATAAGCAGTAAA	ACG
rs4856	CAAGTATTTGAAAATAAG	ACT
rs3782314	ACAAATATATGAGAACTCCTCTTT	ACT
rs2087736	ACTCTGTCTCTACTAAAAATAAAA	ACT
rs2068372	TGGCTTCATGGCACCCTG	ACT
rs2068371	AGCCAAGCCTGTCACTGGCCCT	ACT
rs1184529	TCCTGGTGGTTTGCCACTTA	ACT
rs2101256	GTTTTTTTGGTGTGTTGATATGTA	ACT
rs1552257	AGCTCACCTACGAAAATGAATAA	ACG
rs1565585	TGTCAGACTGCACTACAT	ACT
rs4080800	ATTCTCCTGTCTCAGCCTCC	ACG
rs1388658	GAGAGGATGGGTGAAATAAGG	ACT
rs3782316	TGTAATGTCATTAGGAAGAAACA	ACG
rs1484086	CTCTTCGGAAGTCTCTTTCTCA	ACT
rs1484087	TTTACACCCCCAAATCTAGAG	ACT
rs2291549	CCTCTTTGCTGCCAGTGG	ACT
rs1565584	TGATGCAATAAGTATATATAGTAC	ACT
rs3071166	TAGTACGTGGCTTTTTTTTTTTTT	CGT
rs1907652	TCCAGTACCTAGCACCTAAC	CGT
rs3759119	TTAGTGCCGGGATGAATAACT	ACT

dbSNP rs#	Extend Primer	Term Mix
rs3759120	AAATTGCAACTGTGAGTATTAAAG	ACT
rs3782317	AGAATAACCCACCTGCCAAAT	ACT
rs3825166	ACTTAGCCTACATTTGAAAAGGG	ACT
rs1872191	CTCAGTCCGCTCCCCACTT	ACG
rs8628	TTAAACACTATGACACAT	ACG
rs1872193	AGAGGCAGGACACTAGCC	ACG
rs1042833	CCCTTGGATCTCTTTGAG	ACG
rs2476	GTGACGTTAATGGGACAGCT	ACT
rs1976206	TACAGGCGCCCACCACCA	ACG
rs1388659	CCACATTTGGTAAGTTTTGACAT	ACT
rs1184528	AGAGGTTGCAGAGAGCCAAGATC	ACT

Genetic Analysis

[0228] Allelotyping results from the discovery cohort are shown for cases and controls in Table 13. The allele frequency for the A2 allele is noted in the fifth and sixth columns for osteoarthritis case pools and control pools, respectively, where “AF” is allele frequency. The allele frequency for the A1 allele can be easily calculated by subtracting the A2 allele frequency from 1 (A1 AF = 1-A2 AF). For example, the SNP rs900743 has the following case and control allele frequencies: case A1 (G) = 0.92; case A2 (T) = 0.08; control A1 (G) = 0.90; and control A2 (T) = 0.10, where the nucleotide is provided in paranthesis. Some SNPs are labeled “untyped” because of failed assays.

TABLE 13

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs1058701	230	27004780	A/C			
rs11613	231	27004781	C/G			
rs900743	5330	27009880	G/T	0.08	0.10	0.381
rs2129091	6334	27010884	A/C	0.19	0.20	0.646
rs15556	11372	27015922	C/T	0.74	0.74	0.923
rs1053724	11456	27016006	G/T	0.14	0.15	0.592
rs9699	11501	27016051	A/G	0.67	0.68	0.595
rs4856	13393	27017943	C/T	0.10	0.09	0.474
rs3782314	16666	27021216	A/G	0.20	0.22	0.278
rs2087736	17596	27022146	C/T			
rs2068372	19710	27024260	A/G	0.11	0.10	0.634
rs2068371	19800	27024350	A/G			
rs1184529	20297	27024847	A/G			
rs2101256	20967	27025517	G/T			
rs1552257	32514	27037064	C/T	0.67	0.67	0.969
rs1565585	33159	27037709	A/G	0.69	0.65	0.064
rs4080800	37600	27042150	A/G			
rs1388658	41259	27045809	A/G	0.35	0.38	0.362
rs3782316	41329	27045879	C/T	0.47	0.46	0.740
rs1484086	50060	27054610	C/T			

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs1484087	53292	27057842	C/T	0.35	0.38	0.127
rs2291549	53393	27057943	A/G			
rs1565584	56417	27060967	C/T	0.66	0.69	0.327
rs3071166	56435	27060985	-/TT	0.52	0.47	0.042
rs1907652	58847	27063397	A/T	0.64	0.64	0.984
rs3759119	59595	27064145	C/T	0.08	0.05	0.042
rs3759120	59661	27064211	G/T	0.15	0.14	0.861
rs3782317	60355	27064905	C/T	0.08	0.07	0.780
rs3825166	60407	27064957	A/C	0.83	0.84	0.660
rs1872191	62357	27066907	A/G	0.04	0.05	0.555
rs8628	68230	27072780	G/T	0.59	0.61	0.411
rs1872193	68516	27073066	A/G	0.92	0.91	0.669
rs1042833	69055	27073605	C/T	0.85	0.86	0.605
rs2476	72603	27077153	C/G	0.81	0.82	0.392
rs1976206	73928	27078478	A/G	0.16	0.15	0.781
rs1388659	85897	27090447	C/T	0.49	0.50	0.692
rs1184528	91554	27096104	A/G	0.20	0.17	0.228

[0229] The *TM7SF3* proximal SNPs were also allelotyped in the replication cohorts using the methods described herein and the primers provided in Tables 11 and 12. The replication allelotyping results for replication cohort #1 and replication cohort #2 are provided in Tables 14 and 15, respectively.

TABLE 14

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs1058701	230	27004780	A/C			
rs11613	231	27004781	C/G			
rs900743	5330	27009880	G/T	0.07	0.11	0.258
rs2129091	6334	27010884	A/C	0.18	0.20	0.450
rs15556	11372	27015922	C/T	0.73	0.74	0.858
rs1053724	11456	27016006	G/T	0.15	0.15	0.969
rs9699	11501	27016051	A/G	0.65	0.68	0.290
rs4856	13393	27017943	C/T	0.11	0.10	0.735
rs3782314	16666	27021216	A/G	0.18	0.22	0.171
rs2087736	17596	27022146	C/T			
rs2068372	19710	27024260	A/G	0.11	0.11	0.867
rs2068371	19800	27024350	A/G			
rs1184529	20297	27024847	A/G			
rs2101256	20967	27025517	G/T			
rs1552257	32514	27037064	C/T	0.70	0.66	0.239
rs1565585	33159	27037709	A/G	0.67	0.65	0.599
rs4080800	37600	27042150	A/G			
rs1388658	41259	27045809	A/G	0.36	NA	NA
rs3782316	41329	27045879	C/T	0.45	0.49	0.230
rs1484086	50060	27054610	C/T			
rs1484087	53292	27057842	C/T	0.32	0.40	0.023
rs2291549	53393	27057943	A/G			
rs1565584	56417	27060967	C/T	0.61	0.65	0.297
rs3071166	56435	27060985	-/TT	0.55	0.45	0.003
rs1907652	58847	27063397	A/T	0.62	NA	0.638
rs3759119	59595	27064145	C/T	0.09	0.04	0.025
rs3759120	59661	27064211	G/T	0.16	0.15	0.651

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs3782317	60355	27064905	C/T	0.07	0.07	0.986
rs3825166	60407	27064957	A/C	0.84	0.84	0.945
rs1872191	62357	27066907	A/G	0.04	0.06	0.562
rs8628	68230	27072780	G/T	0.58	0.64	0.060
rs1872193	68516	27073066	A/G	0.92	0.91	0.957
rs1042833	69055	27073605	C/T	0.85	0.86	0.746
rs2476	72603	27077153	C/G	0.81	0.84	0.397
rs1976206	73928	27078478	A/G	0.17	0.16	0.788
rs1388659	85897	27090447	C/T	0.47	0.50	0.283
rs1184528	91554	27096104	A/G	0.21	0.16	0.170

TABLE 15

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs1058701	230	27004780	A/C			
rs11613	231	27004781	C/G			
rs900743	5330	27009880	G/T	0.10	0.09	0.763
rs2129091	6334	27010884	A/C	0.21	0.20	0.896
rs15556	11372	27015922	C/T	0.75	0.75	0.974
rs1053724	11456	27016006	G/T	0.12	0.15	0.373
rs9699	11501	27016051	A/G	0.69	0.67	0.615
rs4856	13393	27017943	C/T	0.10	0.08	0.413
rs3782314	16666	27021216	A/G	0.22	0.22	0.944
rs2087736	17596	27022146	C/T			
rs2068372	19710	27024260	A/G	0.11	0.10	0.561
rs2068371	19800	27024350	A/G			
rs1184529	20297	27024847	A/G			
rs2101256	20967	27025517	G/T			
rs1552257	32514	27037064	C/T	0.63	0.68	0.247
rs1565585	33159	27037709	A/G	0.73	0.66	0.029
rs4080800	37600	27042150	A/G			
rs1388658	41259	27045809	A/G	0.33	0.38	0.217
rs3782316	41329	27045879	C/T	0.51	0.42	0.036
rs1484086	50060	27054610	C/T			
rs1484087	53292	27057842	C/T	0.38	0.36	0.575
rs2291549	53393	27057943	A/G			
rs1565584	56417	27060967	C/T	0.73	0.74	0.635
rs3071166	56435	27060985	-/TT	0.48	0.49	0.727
rs1907652	58847	27063397	A/T	0.66	-0.02	
rs3759119	59595	27064145	C/T	0.06	0.05	0.669
rs3759120	59661	27064211	G/T	0.13	0.14	0.881
rs3782317	60355	27064905	C/T	0.08	0.07	0.655
rs3825166	60407	27064957	A/C	0.82	0.84	0.546
rs1872191	62357	27066907	A/G	0.04	0.04	0.894
rs8628	68230	27072780	G/T	0.61	0.56	0.234
rs1872193	68516	27073066	A/G	0.93	0.91	0.528
rs1042833	69055	27073605	C/T	0.85	0.86	0.682
rs2476	72603	27077153	C/G	0.80	0.81	0.829
rs1976206	73928	27078478	A/G	0.14	0.14	0.834
rs1388659	85897	27090447	C/T	0.51	0.48	0.499
rs1184528	91554	27096104	A/G	0.18	0.18	0.993

[0230] Allelotyping results were considered particularly significant with a calculated p-value of less than or equal to 0.05 for allelotyping results. These values are indicated in bold. The allelotyping p-values

were plotted in Figure 1 for the discovery cohort. The position of each SNP on the chromosome is presented on the x-axis. The y-axis provides the negative logarithm (base 10) of the p-value comparing the estimated allele in the case group to that of the control group. The minor allele frequency of the control group for each SNP designated by an X or other symbol on the graphs in Figure 1 can be determined by consulting Table 13. For example, the left-most X on the left graph is at position 27004780. By proceeding down the Table from top to bottom and across the graphs from left to right the allele frequency associated with each symbol shown can be determined.

[0231] To aid the interpretation, multiple lines have been added to the graph. The broken horizontal lines are drawn at two common significance levels, 0.05 and 0.01. The vertical broken lines are drawn every 20kb to assist in the interpretation of distances between SNPs. Two other lines are drawn to expose linear trends in the association of SNPs to the disease. The light gray line (or generally bottom-most curve) is a nonlinear smoother through the data points on the graph using a local polynomial regression method (W.S. Cleveland, E. Grosse and W.M. Shyu (1992) Local regression models. Chapter 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.). The black line provides a local test for excess statistical significance to identify regions of association. This was created by use of a 10kb sliding window with 1kb step sizes. Within each window, a chi-square goodness of fit test was applied to compare the proportion of SNPs that were significant at a test wise level of 0.01, to the proportion that would be expected by chance alone (0.05 for the methods used here). Resulting p-values that were less than 10^{-8} were truncated at that value.

[0232] The exons and introns of the genes in the covered region are plotted below each graph at the appropriate chromosomal positions. The gene boundary is indicated by the broken horizontal line. The exon positions are shown as thick, unbroken bars. An arrow is placed at the 3' end of each gene to show the direction of transcription.

Example 5

In Vitro Production of Target Polypeptides

[0233] cDNA is cloned into a pIVEX 2.3-MCS vector (Roche Biochem) using a directional cloning method. A cDNA insert is prepared using PCR with forward and reverse primers having 5' restriction site tags (in frame) and 5-6 additional nucleotides in addition to 3' gene-specific portions, the latter of which is typically about twenty to about twenty-five base pairs in length. A Sal I restriction site is introduced by the forward primer and a Sma I restriction site is introduced by the reverse primer. The ends of PCR products are cut with the corresponding restriction enzymes (*i.e.*, Sal I and Sma I) and the products are gel-purified. The pIVEX 2.3-MCS vector is linearized using the same restriction enzymes, and the fragment with the correct sized fragment is isolated by gel-purification. Purified PCR product is ligated into the linearized pIVEX 2.3-MCS vector and *E. coli* cells transformed for plasmid

amplification. The newly constructed expression vector is verified by restriction mapping and used for protein production.

[0234] *E. coli* lysate is reconstituted with 0.25 ml of Reconstitution Buffer, the Reaction Mix is reconstituted with 0.8 ml of Reconstitution Buffer; the Feeding Mix is reconstituted with 10.5 ml of Reconstitution Buffer; and the Energy Mix is reconstituted with 0.6 ml of Reconstitution Buffer. 0.5 ml of the Energy Mix was added to the Feeding Mix to obtain the Feeding Solution. 0.75 ml of Reaction Mix, 50 µl of Energy Mix, and 10 µg of the template DNA is added to the *E. coli* lysate.

[0235] Using the reaction device (Roche Biochem), 1 ml of the Reaction Solution is loaded into the reaction compartment. The reaction device is turned upside-down and 10 ml of the Feeding Solution is loaded into the feeding compartment. All lids are closed and the reaction device is loaded into the RTS500 instrument. The instrument is run at 30°C for 24 hours with a stir bar speed of 150 rpm. The pIVEX 2.3 MCS vector includes a nucleotide sequence that encodes six consecutive histidine amino acids on the C-terminal end of the target polypeptide for the purpose of protein purification. Target polypeptide is purified by contacting the contents of reaction device with resin modified with Ni²⁺ ions. Target polypeptide is eluted from the resin with a solution containing free Ni²⁺ ions.

Example 6

Cellular Production of Target Polypeptides

[0236] Nucleic acids are cloned into DNA plasmids having phage recombination sites and target polypeptides are expressed therefrom in a variety of host cells. Alpha phage genomic DNA contains short sequences known as attP sites, and *E. coli* genomic DNA contains unique, short sequences known as attB sites. These regions share homology, allowing for integration of phage DNA into *E. coli* via directional, site-specific recombination using the phage protein Int and the *E. coli* protein IHF. Integration produces two new att sites, L and R, which flank the inserted prophage DNA. Phage excision from *E. coli* genomic DNA can also be accomplished using these two proteins with the addition of a second phage protein, Xis. DNA vectors have been produced where the integration/excision process is modified to allow for the directional integration or excision of a target DNA fragment into a backbone vector in a rapid *in vitro* reaction (Gateway™ Technology (Invitrogen, Inc.)).

[0237] A first step is to transfer the nucleic acid insert into a shuttle vector that contains attL sites surrounding the negative selection gene, ccdB (e.g. pENTER vector, Invitrogen, Inc.). This transfer process is accomplished by digesting the nucleic acid from a DNA vector used for sequencing, and to ligate it into the multicloning site of the shuttle vector, which will place it between the two attL sites while removing the negative selection gene ccdB. A second method is to amplify the nucleic acid by the polymerase chain reaction (PCR) with primers containing attB sites. The amplified fragment then is

integrated into the shuttle vector using Int and IHF. A third method is to utilize a topoisomerase-mediated process, in which the nucleic acid is amplified via PCR using gene-specific primers with the 5' upstream primer containing an additional CACC sequence (*e.g.*, TOPO® expression kit (Invitrogen, Inc.)). In conjunction with Topoisomerase I, the PCR amplified fragment can be cloned into the shuttle vector via the attL sites in the correct orientation.

[0238] Once the nucleic acid is transferred into the shuttle vector, it can be cloned into an expression vector having attR sites. Several vectors containing attR sites for expression of target polypeptide as a native polypeptide, N-fusion polypeptide, and C-fusion polypeptides are commercially available (*e.g.*, pDEST (Invitrogen, Inc.)), and any vector can be converted into an expression vector for receiving a nucleic acid from the shuttle vector by introducing an insert having an attR site flanked by an antibiotic resistant gene for selection using the standard methods described above. Transfer of the nucleic acid from the shuttle vector is accomplished by directional recombination using Int, IHF, and Xis (LR clonase). Then the desired sequence can be transferred to an expression vector by carrying out a one hour incubation at room temperature with Int, IHF, and Xis, a ten minute incubation at 37°C with proteinase K, transforming bacteria and allowing expression for one hour, and then plating on selective media. Generally, 90% cloning efficiency is achieved by this method. Examples of expression vectors are pDEST 14 bacterial expression vector with att7 promoter, pDEST 15 bacterial expression vector with a T7 promoter and a N-terminal GST tag, pDEST 17 bacterial vector with a T7 promoter and a N-terminal polyhistidine affinity tag, and pDEST 12.2 mammalian expression vector with a CMV promoter and neo resistance gene. These expression vectors or others like them are transformed or transfected into cells for expression of the target polypeptide or polypeptide variants. These expression vectors are often transfected, for example, into murine-transformed adipocyte cell line 3T3-L1, (ATCC), human embryonic kidney cell line 293, and rat cardiomyocyte cell line H9C2.

Nucleotide and Amino Acid Sequence Embodiments

[0239] Table A includes information pertaining to the incident polymorphic variant associated with osteoarthritis identified herein. Public information pertaining to the polymorphism and the genomic sequence that includes the polymorphism are indicated. The genomic sequences identified in Table A may be accessed at the http address www.ncbi.nih.gov/entrez/query.fcgi, for example, by using the publicly available SNP reference number (*e.g.*, rs1484086). The chromosome position refers to the position of the SNP within NCBI's Genome Build 34, which may be accessed at the following http address: www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=hum_chr.inf&query=. The "Contig Position" provided in Table A corresponds to a nucleotide position set forth in the contig sequence (see "Contig Accession No."), and designates the polymorphic site corresponding to the SNP reference

number. The sequence containing the polymorphisms also may be referenced by the "Nucleotide Accession No." set forth in Table A. The "Sequence Identification" corresponds to cDNA sequence that encodes associated target polypeptides (*e.g.*, TM7SF3) of the invention. The position of the SNP within the cDNA sequence is provided in the "Sequence Position" column of Table A. If the SNP falls within an exon, the corresponding amino acid position (and amino acid change, if applicable) is provided as well. Also, the allelic variation at the polymorphic site and the allelic variant identified as associated with osteoarthritis is specified in Table A. All nucleotide and polypeptide sequences referenced and accessed by the parameters set forth in Table A are incorporated herein by reference.

Table A

RS_ID	Chromosome	Chrom Position	Contig Accession No. [1]	Contig Position	Nucleotide Accession No. [2]	Sequence Position	Amino Acid Position	Locus	Locus ID	A [3]	Allelic Variability	OA Assoc. Allele
rs1484086	12	27054610	Hs12_9871_34:16	19922317	NM_016551	intron		TM7SF3	51768	R	[T/C]	T

[1] Contig Accession Number which can be found in the NCBI Database:
[http address: www.ncbi.nih.gov/entrez/query.fcgi](http://www.ncbi.nih.gov/entrez/query.fcgi)

[2] Sequence Identification or Nucleotide Accession Number which can be found in the NCBI Database:
[http address: www.ncbi.nih.gov/entrez/query.fcgi](http://www.ncbi.nih.gov/entrez/query.fcgi)

[3] "A" column is the sequence orientation ("F" is forward, "R" is reverse).

[0240] The following is a genomic nucleotide sequence for an *TM7SF3* region. The following nucleotide representations are used throughout: "A" or "a" is adenosine, adenine, or adenylic acid; "C" or "c" is cytidine, cytosine, or cytidylic acid; "G" or "g" is guanosine, guanine, or guanylic acid; "T" or "t" is thymidine, thymine, or thymidylic acid; and "I" or "i" is inosine, hypoxanthine, or inosinic acid. Exons are indicated in italicized lower case type, introns are depicted in normal text lower case type, and polymorphic sites are depicted in bold upper case type. SNPs are designated by the following convention: "R" represents A or G, "M" represents A or C; "W" represents A or T; "Y" represents C or T; "S" represents C or G; "K" represents G or T; "V" represents A, C or G; "H" represents A, C, or T; "D" represents A, G, or T; "B" represents C, G, or T; and "N" represents A, G, C, or T.

TM7SF3 Genomic Sequence (SEQ ID NO: 1)

>12:27004551-27096350

```

1      acccaaatac taattttcat gttatttttg aacagatgcc tttttggaaa tcttaaaaaa
61      gcaaacattt caaacagttg gctaacacgc atgaactaac actcattcag tttttaaaatt
121     acggatgata tccgtattaa cattataatg actctatact acagattgac atggtacatc
181     gtaacaagtc cgaaggattc ttccttgatg catctcgaca catccttgam Scacctcaac
241     atggactgga gagaaggcac ttggaagcaa atcagaatgt aactaaata aacagtcaac
301     ttttggggtg tggatggaag gggggtccat tttaaaagtg cttttacatt gaatttcctt

```

361	cccagattag	atcagcaaat	aaatgaaatt	tattaaaaga	atgttggtgc	tttaatagcc
421	atgctgttca	gtgtagaaac	tttaaatctc	tcattaattt	tctgtgtatt	gtatcatgta
481	cagcctgtta	gatatattgca	ggttttatag	gttattcagt	ccaacgcttt	atatttttgc
541	tcttggttagt	tgttgctgtg	aaacagtcac	ttcctttctg	gacatcaaag	atattttctt
601	ggcattaatt	ttaaactgat	ttcataaggt	caaagtatta	aatgttttaa	ttcagattgt
661	ggatagtttg	tagtcctttg	attttgcaaa	tgaaattaaa	atgcctttgt	ttagatgta
721	ggattaagta	tagtagtgct	ctctccattg	tgtcttaatt	tttttttaac	tcacttggtg
781	caggtccttta	ttgaataaag	gaggaataaa	cctagacagg	gagttagaaa	tgaaagacat
841	ttcaagacag	ttcaagggtt	atgttggtga	cacttcttgt	cttccagaca	gttttgaaag
901	cacagatcct	gcctcttgtg	cttggttaaaa	aatctgtgat	acttctctga	agattttata
961	attaaataag	tgaaccatta	ccatgacatt	ttataaaagt	gctgccctat	ttgaagtatt
1021	gggttttttc	tcatgatgag	cttggttaga	cttggtactg	aaaaaatcag	tagttttttt
1081	ccttctatct	taaaggcaaa	gaaatgcttg	cttcagctag	gcatggtggc	tcatgcctgt
1141	aattgggagg	ctgaggtggg	tggatcacct	gacgtcagga	attggagacc	agcctggcca
1201	acatggtgaa	accccatctc	tactaaaaat	acaaaaagt	agccaggcgt	gatagcgggt
1261	gcctgtaatc	ccagctgctc	gggaggctaa	agcaagagaa	ttgtttgaac	ctgggagtg
1321	gaggttgacg	tgagccgagg	tcgtgccatt	gcactccagc	ctggccaaca	agagcgagac
1381	tctgtctaaa	aaaaaagaaa	tgcttgcttt	ctgtgtagtt	ttttaactta	ggtcagtgcc
1441	ataggactga	ccaaatcata	aattgaggat	gctgttggtg	ttcaaagt	ttactcatta
1501	cccttactca	gtttgcagct	gcaatgatct	gtacttaacc	tgcttttaca	tattatgctt
1561	ccttgacaat	gggaggaatt	ttaatgggta	acttgagaa	tatgacagat	acttgcttta
1621	atctacttct	tcctaccata	atctaattca	aaactatcca	gtagaacttt	ctgaaatgat
1681	tggaatgttt	tttatctgtg	ctgtccagtg	tggaagccag	tagccacatg	tggctgttga
1741	gcacttgacg	tgctcactaat	gcgtttgagg	gactgaattg	ttaattttta	aaaatgtatt
1801	gatttttaatt	aagtttaaca	ttgttttgga	cagcacagat	tagcctttta	gagtaccttc
1861	ttgtctgtag	ctagcagtg	ataggtttcc	ccatgaaaag	gacctacatt	ttgttggtt
1921	aagagggtat	taaatgtgct	ttgccaagaa	tactaccttg	tcatttcttc	cagcttatta
1981	gttccgtggg	tttcttgagg	tactgttaaa	tcctttttac	aaatatatat	atacacaact
2041	actgaagaga	ctgtttatct	ggggagttaa	tcacttttat	atttaagtct	taggcttaat
2101	atgataaaat	acctctcagt	tgctgtgaat	aatggtctat	agaagctgtt	catttgagaa
2161	atgattattc	tcgagtttct	gttgatttta	aacctttttt	gtctgtctct	ttcactagac
2221	tgtaagcttc	ttcagatcag	gggacatgtc	ttgacagtct	ttgtattcct	gttggttagc
2281	atagtactgg	cacatcatag	gtgtgcaaga	taagaaaatg	agattggaag	aatttatgct
2341	acttgctgag	cctaacaaat	gtggatgtct	gtattatact	gatactatat	tttcaatagt
2401	tagcaaagaa	ggccatttta	tatttttaata	ctgaaaagca	gaattgtgct	catttttatc
2461	aatacttacc	atataccata	cactagatag	taggtggatg	taaagatgca	gttctgtctc
2521	tcaaagagga	ttatgaccta	gaaaagggtta	cagattggat	tgagataact	ttataggcct
2581	gaacaatata	gaaaagattt	ctaaatcatt	gaagacttca	gggaggaaaa	gagattgaat
2641	gcgtcggcaa	tttttatcat	tcccggtact	ttgaacaata	gtgatgttaa	gtgcaaaaaa
2701	aaagaaaaag	ttgtctgatca	ttgtggtgag	aggaatgtaa	aatatgagta	tttttcttgc
2761	tgctaggttt	agttatatgt	tgctgcttat	tcttaattct	agtcctaaca	aatttgattt
2821	caaaagcaca	tctgtggtca	accaacttca	ttgttaatat	ttgttttatt	catattaagt
2881	gacaatgtaa	gaatattcat	gttaagtgg	actgcagatt	ttataaaaa	gagaggatcc
2941	aaaaataaaa	aagtattttg	ctgattttcca	actcccatcc	ctctttttca	ggaactgcaa
3001	gcacatgttg	accagataac	tgaaatggca	gcagtaatga	ggaaagccat	tgaaattgac
3061	gagcaacagg	gttgcaagga	acaagaacga	atatttcaac	ttgaagtaag	ttttacattg
3121	caaatgtaga	tgaaaactgt	ccatataaag	atttatgcca	gaaaatatat	cccattttta
3181	aatatatatt	tactcttctt	gggtactgg	aatgttctcc	tggaatgctg	ctatgtgggc
3241	atattatgca	agaccattta	tgtggtgatg	caagaatcta	ggattattag	taacttacaa
3301	tataccagta	atgtagaaaa	ttgatgtttc	attctatatt	gcttatctta	tgaagtgaat
3361	gtacttcttc	aattgagtca	tttttaaaat	tttgtttatt	ctattagaga	tcatagtgat
3421	acagggaaat	tttttttaat	ccacaaatat	ttaggggcac	tcttattttt	ctggggattc
3481	taaaaacctt	tgctcatctg	cctttatttt	atgacagtga	aagtaccagt	ctctgaataa
3541	ctgaaatcag	tacaatagac	ttttatgcct	taagagcttc	attttgaat	cacggcgggg
3601	aacctgcttt	agactactaa	aatttagaaa	gtgagactga	agataaagt	atagagaaca
3661	aatttgaaaa	gggtattggt	aaatgtctac	aaattttttt	acagtaacta	accaaaggga
3721	ggccagaaag	caaagtattt	taccagtgcc	agatttggtc	acttactttc	cttcttgaaa
3781	aatatcatca	gctaccctta	cctgtctgct	ctcaaagcc	aatgagctaa	acttgcctta
3841	cattccatgc	tatcatggaa	agtaataaaa	atactaaaaa	ttactaggag	attctagcct
3901	attcactggc	tttacgagtt	tatatataaa	taaactctaga	tgttattttc	ttcatctcta
3961	atgtgattgt	aaaataaggc	aatttctgag	cttctgttta	caattatagc	tgcttaagta
4021	agaaggcttt	gcagtggtgg	aagatgttga	acacgattcc	tagctgcata	agtgttcttt
4081	agattcaagt	tgccaactgc	atatatcatt	gagtagatta	atgccaatta	cagaccttgc
4141	ttgtttgtta	taccttcata	tgttcagttt	ttcctaaca	atgtttgtga	tgttgtagct

4201	accttgtaga	tagtgactgc	atcctttgtt	ttgcttgcc	tttagcaaga	aaacaaaggc
4261	ttgagagaga	tccttcaaat	aactcgagaa	tcatttttga	acctaaggaa	agatgatgcg
4321	tcggaagagta	cttctttgtc	agcatttagtg	accaacagtg	atttgagtct	gaggaagagc
4381	tgaagagttt	ctgagtgctgt	gagcttctta	catggctcca	aatggctcaa	taagtgaatg
4441	aatgaatgga	cagaaaaatc	aatcctttat	ttttttctct	gtaaatatgt	acagtgacag
4501	ggctatgaga	tagcaacaaa	aaatgcatag	ttaatgggtca	tagactttat	tccaaaacat
4561	aattggaaaa	tagaaactga	gccattgcca	aatggtaaag	aaatgaaaag	ttttcacagt
4621	gactactgaa	tataccaaga	gcttttgcca	gtactgctgg	ctttctgggt	gattaattag
4681	gtaaaccttga	atattcccag	taaatgtttg	agaatgcata	aaattatacc	aatttaaaaa
4741	tataaatttt	tgccactggt	ttgtgaacgg	aggatttgta	cgctaaattt	catccttatt
4801	tggcgtaaaa	gttgatgcaa	agtttatgta	ccatgagact	tctagtgtt	gcctactttc
4861	acattttatt	aaatttttgt	gaaagaaaaa	acagttttta	aggctatagg	aatgtgttat
4921	atgtacctta	aacagataat	ttaatttttag	caatttatca	taactgatac	ttagctaaca
4981	attgtccata	acttggtttta	gtgtaagaaa	taccatttca	ggaatgaaaa	agaaggaata
5041	ctactttcta	agaaagaaga	tcttataata	gacataattt	gtaggttaaa	ctactccttt
5101	gaaagaataa	gttttggttc	aaagtcaata	atgaagacga	gatttggttt	tctcttctct
5161	ggctgatttt	taaggatatt	atgtagtcca	tttagttaac	agaattgaaa	tgtttgatat
5221	ggcaagatag	gtatggtaat	ttcaaagtga	attgggaatt	cctctggctc	atagaaccct
5281	tttttttttc	ctttaagtat	tcttgagata	caaaaaaaaa	aagtaaaata	aatttcaaaa
5341	aaaaagttcc	ggatctgttt	ttaagctcca	tctggctctc	ataacctgca	agatttttct
5401	taaaaccttt	cagctgaaag	tgggggtaaa	ggtggagtaa	tctgtggatt	tgtttctgtt
5461	gtctttttaa	atgtcaaata	tataatatgt	aattttttta	aaaaccacca	gatacagaaa
5521	tgtgctttta	catcagttga	aacctaaatt	ttcttatgtt	gtggtgattg	tattaaaaag
5581	ggataaaaaga	agagtgctca	acatgggttaa	atatattgta	ctcatttatg	ttgaatacgt
5641	atataaaatta	agacaaatgg	aaaattatac	tttgagtata	taatttggtt	aatattactt
5701	tatatggtaa	ttttatgtat	aatttcatat	attggtaaaa	ttcaaaaacta	cacttgagaa
5761	tttttttatc	ttaagtttgg	ggtgaatggg	gtggatgaga	ctgattgaat	agaaaagggc
5821	taatggccca	aacatttat	agatttcttt	ttttcagtca	gaggccttat	ttgatatttt
5881	ataaataaat	gacagttttt	attttttaaac	tttttatgtt	ttttgggaaa	gtatttcctta
5941	atttaatgac	acattcattc	agatacttct	tatccctgct	aataaaggaa	atctatttca
6001	agctacacca	ttgagattaa	gtctgagcca	gttcattgag	gcagctctac	tataaaagct
6061	tacttgataa	ataattattt	ttgtaaacaa	gttgggttaa	cttattcttc	gtctttttgc
6121	ttggatatga	atttaagggtc	ttcatgttta	aagacattta	ctttgttatt	tagtgacaca
6181	tttccatcct	attttttttt	tttttttggt	tgttgttaaa	cagaacctta	agtttatgtt
6241	tggaggtatgt	actgcatagg	aacctatttt	attattaaag	atgaatgatt	aaaattggta
6301	tggtctccaa	tttaatttga	aaagtgccta	cccKtattct	tatatatggt	tttaattttta
6361	ggttttttgt	ctcttcttag	tgcaaaacta	cttagcagtg	acctctatct	gtatttcctta
6421	ggaattagca	gcttcttagt	gtggatcctg	cagaacttct	taccatttgt	agtaggttga
6481	atcatgtccc	ctagaaggta	agtctaagtc	ctaacttgat	acacctggga	agggtgacct
6541	atttggaat	agtcctttaca	gatgtgatta	gggatctca	ggatataatt	accttggtat
6601	ttgggtgggc	tccaaaccca	gtgactagt	tcgttacaat	aagagaggag	gagatatgaa
6661	acacagatat	agcagcatga	ggacagaggc	aatgagttac	gcaaggactg	ctagccacca
6721	ccagaagctg	caagagaggc	atggaatggg	ttcttcacgg	gctccagaaa	ggaccaagcc
6781	tgcttgattt	cagacttctg	gctccagaaa	ctgtgaggaa	ataaatttct	gttggtttca
6841	ttcaccagc	ctgtggtaat	ttgttagagt	aaccttagga	aactaagaca	tctattgttt
6901	ttcatctatc	aactatattt	ttattttata	taagaaagtt	cttaaaaacc	aatttataac
6961	ctcacaattt	ctgctcta	tttatattct	ttaatatcaa	ggtgacttct	aaaaattttat
7021	agaaataaat	tgcttccctg	taatcccagc	actttgggag	gccgaaacag	gtggatcagt
7081	tgagggtcagg	agttcaagac	cagcctagcc	aacatggtaa	acctgtctc	tactaaaaat
7141	acaaaaagtt	agccagggat	ggtgggtggg	gcctgtaatc	tcagctactt	gggagccga
7201	ggcaggagaa	tcgcttgaac	ctgggaagca	gaggttgag	tgagccgaga	tctcaccatt
7261	gcactccagc	ctgggcaata	atagtgaac	tccatctcaa	aaaaaaaaaa	aaaattgctt
7321	caatgaaaaa	tggaacaatc	tagatccata	attttgggtg	tcactacctt	ttttcattgt
7381	cagagtga	taggagaaca	tttttttagta	aaatactgat	gaagatgata	atttctaatt
7441	ggaatttgag	gagacacttt	ttcctttaac	ctctctggct	gctcctcagt	ctcctttgtt
7501	tcttctcgct	gacctcttaa	cattagagag	ccactgggct	tagtccttat	acctattctg
7561	tgacgtgtct	ctactcactg	gtgatctcac	ccagtctcat	gactaaatac	cacctctatg
7621	cttgatggct	cctaaattca	cagttccagc	atagacctct	caccccggt	tcatgcagct
7681	gtgactgtca	aagagaaatt	ttatgcttaa	gatgtccaaa	gccaaactcc	aaacctccca
7741	ggacgcctcc	atactctttc	ccatcctgct	tgatgacaac	tccatccttg	cagttgtcaa
7801	ggccaaaaac	cttgggtggca	actttgactc	cactttctca	tatcgtgcat	cctattcatc
7861	agcaaatcct	gttggtggct	cttctctcca	aataaatcag	aatcctacca	tttctctcct
7921	ccactgctac	ccctgggtcaa	agctacatta	tggtagtgtt	aaagtataca	aattggattc
7981	aacctaaata	tcctacagta	ggggaatagt	ttaatgtcaa	tattcactca	atacagttga

8041	ttacagtcta	tggcttgaag	atgctagcaa	caagaagaaa	aaagtaccta	aatatgcatt
8101	aaaaagctta	acataaatta	cccatataat	tacaaatatg	taaaaatcag	atgcatacgg
8161	ggaaatgtcc	tgaagcaaac	agcatcagta	ttcttccagc	taattgtact	gtttgcatta
8221	gccagtcagt	ttctgatgcc	tgtaaatcaa	gcatttttgt	agctaccatt	atggagttgt
8281	aattttattca	tcagttcccc	caactgatgg	atagtttgca	aagtttaata	caaaaataatg
8341	cagtaaacat	ccatatacat	aactttat	attgtaattt	ctagcgtgca	aattcccaga
8401	agtttgtag	gtcatgtttt	gttatttttt	aaatgaaaaa	ataaatgtac	actgttaaaa
8461	atcttagcat	actaaaaaat	ttacaattca	gagtaaacct	cttcctacaa	cctctgtctt
8521	tccagagggtc	attataagtt	tcttgatgaat	cagaaaaac	gttaaccaaa	aaaagtgact
8581	gaggcaagt	tcttaaatca	gtagagggtt	attgggccaa	agtttaaga	catgccagag
8641	aaaaaacgca	agccacatga	gcattctgtga	cctgagcttt	ctaaagagg	ttttggaaac
8701	tcattatttta	ggggagagag	tgaggaggga	aaagggagag	gggttaggca	gtgaggcaaa
8761	tgattacatt	cttgtgaggc	tctgattaac	gctcagtaaa	tctatgtttc	atataagata
8821	aactgaaaaa	agggagtaaa	gactcagtta	ttcatctgtt	tcagggtagg	cagaagagtg
8881	gttgattttg	tctcatcctt	attctgtacc	tgggaagata	agcttggtat	cagcattgtc
8941	agtgtgagat	ttacaaaaac	ttggttttaa	gagttaaact	taggttgag	acccaaggtt
9001	acaattggca	gtgtgtgtt	ttacagctct	ctacaatgag	aaggggaggc	cttgaggctt
9061	ctgttaataa	tactccccct	cctacacatg	acatacttgg	ttttggctac	tcctactcaa
9121	cctccaagtt	tcagctggaa	tgtcacttcc	tcaaggactc	cttccttggc	atcccacttt
9181	ctagccctat	cccattatac	attttcattg	catattacac	atcttgtgta	actattttag
9241	tatctgtctc	acactccagt	cccagacagc	aaagaccatg	cctgtgtttt	ccatcttctc
9301	agtgtctcaag	gcaatgcctg	gccagggaact	caacaaatat	ttaatgagtg	aatgtgagat
9361	ctaagtcaaa	gcctcactat	catatatata	taatttttat	ttttttgaga	cagtgtcttg
9421	ctttgtcgcc	taggctggag	tgcaatagtg	cgatctcagt	tcactgcagc	ctccgcccc
9481	cgggttcaag	tgattcttcc	tcctcagcct	cccaagtagc	tgacaggtat	gtgccatcat
9541	gcccggctcg	ttatttttat	tttttagtag	agacggggtt	tcaccatgtt	ggccaggctg
9601	gtcttgaaact	ctgacttca	agtgtactgc	ccaccttggc	ctctcaaagt	gtcgggatta
9661	caggcgtgag	ccaccacgcc	actcctgtat	aacaattttt	taaaaaatct	taagcttgat
9721	ccagacctgc	caacctctcc	agtgtaatga	ttacatggta	gtccctgggt	aagtccagag
9781	ttcccactaa	cttaagcccc	tctttgtgtc	aatttcataa	agagtgggtc	ttgcaaatac
9841	aatggccaaa	ttggccaagag	gcatttcaat	caacagtcag	ggtaatgtgt	gattctttga
9901	cataacagaa	ttccactgac	tgaatggctt	aaacaacaga	aataatttct	cacagttctg
9961	gaggctgaag	tccatgatca	agggtcagca	gggatgggtt	cctctgcttg	cagatgctcc
10021	cctcttactg	ctcttcacac	ggtctttccc	ctgtgagcac	atgcacctgg	tgtttctgtg
10081	tcctaattctc	cccttctttac	aaagacactg	gtcagactgg	attaggaccc	accctaatag
10141	tctcaatttta	atttgaacc	tctttaaagg	ccccgtctcc	aaataaggtc	acattctgag
10201	gtaccgggtg	ttagggtctc	aacatataaa	tttgtttggg	gggcacagtt	tagtctatca
10261	gcaaaccata	ggtagctgcc	ctcccaacaa	gggctgggat	gcaatgaggg	gctgtgccct
10321	agggcaaggc	cctgttctca	ggaaattaga	accaagaact	ctgggtttga	atgaatgtta
10381	ctaacaggta	ttaagtacat	cacctaggag	tgaacttcgg	accaagtttc	tattgtggag
10441	atatggatat	tggaaatggt	aaatattatg	taccaatcat	ttgaaataaa	tgggtccatta
10501	cttcagataa	ttaaacttga	aaaacatgct	ttcactgtgt	aaaattaata	attttcaaaa
10561	agaacgatcg	tagcaatact	tatattagaa	acatcctatc	aattcgtag	cagaagaaaa
10621	tagaaaatac	ctgggttttac	atctgtatca	tgatgagtg	aaagcacatt	tggataggga
10681	gtcaggcatc	ccagtttttg	tctgagcaat	gttagttgat	aactctctga	cttcatatca
10741	ttaaacttaag	aatttctctca	ggtatttaaag	aggtcagtga	acatgacctc	tgaagtcctt
10801	tgaaggttta	atattctagc	aatttctttg	ttggaatcag	ttttctattt	cacctgctac
10861	agaaaactca	actggcctgt	ggtacaatat	acctagtag	gtttggcaac	ctggtagcta
10921	acctataaca	gattaatcag	aactttctga	aaacatttta	aagaacagtt	caacagagtt
10981	tcctgttaac	tactactttt	cttttagctg	acacacctga	agtgctaata	acaccaaatt
11041	agtcttttaa	cacattgccc	cagatttcat	ttcctgttat	tatctttgtg	atttttttaa
11101	aatgaagaac	tatattgttt	gcccataata	aataactttt	atagctatga	aactcgatcat
11161	gtcttgtttg	tgtggtacaa	atataatctg	aaaaatttta	tgaataattt	taatgaggct
11221	tttttagcatt	tttttcaaaa	tagggattta	ttataaaaac	tatacaaaact	ttttatcact
11281	cacaaatgat	gaaaaatttt	aaacaattag	tatcttccct	ccaaaatagt	gtgttctgtc
11341	ttttagctct	tcactggcag	tgcattcaga	tRgtatagaa	aataacaaaac	atttcataata
11401	actttgtccc	tacagtaaat	aattttttta	aacttttcat	caagtttcag	tgtaMtga
11461	gttaattata	gacagtaagg	attacagaat	aagcagtaaa	Yggcacctg	cactgtattt
11521	ttcaaagcaa	aaattgtttt	atattttcta	tgttgtttta	tcctgcccc	aacacttaaa
11581	gcaaaagtta	aatatttttt	cagttacctg	aaacttcaaa	aatctattgt	actttgaaaa
11641	aacttttaaaa	agcagcatca	tttaaatagt	ggttaactca	taaactgtga	tgcaacaaca
11701	ttccacaacc	ctaataccac	tcaggcctgc	agagtacca	gctcacaata	attcagtgtc
11761	ctgaaagctg	gggggaaagc	ccctgagtag	taagtatgca	acccagggtg	tcgggtgctt
11821	tcttgtggtg	aaacaagggt	gtcaaaactg	aggggactct	ccaaggtcct	aatgaaaagc

11881	agctcctcta	aacacctgaa	tcctttttga	ctccattctc	atcactttaa	ggatgaagcc
11941	tttctgtaca	aaggctttgt	acagaaatta	ttaaaaaacta	tattttacac	taatgctaac
12001	agctaattga	gcagaaagaa	ctatgaagta	caaatttttct	taaggttttc	ttattaatgg
12061	caatacagac	tttttttttt	tttgagacgg	agtctcgcctc	tggtgcccag	gggtggagtgc
12121	agtggcgtga	tcttggtctc	ctgcaacctc	tgcctcctgg	gttcaagcga	ttctcctgcc
12181	tcagcctccc	gagtagctgg	gattacaggg	acctgccaca	ataccagct	aatttttcta
12241	tttttagtag	agacaggggtg	tagtagtata	tatttacta	tactatagta	gagaccacta
12301	tggtggccag	gctggtctcg	aactcctgac	cccaggtgat	ccaccgcct	cgacctccca
12361	aaatgctggg	attacaggtt	tgagccaccg	cacctggtcg	gcaatataga	ttaagggtggg
12421	ggttattttcc	taaaaaacaaa	agtaatccat	atgtcatcta	atgcagttaa	attaaagatc
12481	acctgttgca	actatttgaa	aaggccttga	aaaaaaaaaag	ggggccatta	tttggaaat
12541	aaatcacctc	tataaaattc	tattatctat	ataaattggc	ctattttcaa	ataataattt
12601	agctattttca	attaccattt	taagcaataa	aaggcaaaaa	caaagggaa	ctactgaatt
12661	gtacacttaa	acatgaataa	gaaggtaaat	ttcatgtgta	ttttatcaca	acttaaaaaat
12721	tttaaaaaat	actgaaaaaa	aaacaaaggg	taacttttgg	ataatgtatg	ataataagaa
12781	gggcatgcaa	ctaagtaatc	agactgcttt	tagtaagaaa	ataagttttt	tttttttcta
12841	atttctcgag	acagggctct	tgtcacccaa	gctggagtgc	agtgcacaaa	tcaaggctca
12901	ctgtagcctc	aatcttcagg	gctccaggga	tctctccatc	tcagtctcct	tgggagctgg
12961	gagtaggcat	gtgccaccat	gcctggctaa	ttttttaatt	tttttgtaga	gatggggctc
13021	tgtcatgttg	cccatgtcgg	tctcaaaact	ctgggctcaa	gcgatactcc	caccttggtc
13081	tccagttatt	gggattacag	gtgtgagcca	ccatgtctgg	cttgcttctc	tttttgattt
13141	ctaaaaattca	aaggcctaag	tatcaaatcc	ctaaatctcc	aaatactgtc	acagataaag
13201	actcaataat	aaactccctc	cgaaagttaa	gacaggctca	ggtgagagac	ttgtttcaag
13261	gggttatataa	aagaaacacc	agtgtctctg	agaagaatca	gtttttaatt	tttttaattg
13321	atctattttta	tgaataaagt	tgatcataga	tttgtaaacc	aaaaggtaat	ttctcaagta
13381	tttggaataa	agRaaaagcc	tccctaccac	caaccttttg	gtcatctttc	tcatttctgt
13441	acaatcatcc	taatccccta	gtacaccctt	accatataatc	aataagggca	ccataatatt
13501	atgcaaagaa	cagatatata	tgcctgatct	cttattagac	ttgcaccaga	gactgttgaa
13561	ccactccagg	catgaactcc	aaagctgagg	cacactgacc	aagcccctgg	gcatctacag
13621	aagcaaaggc	gttctctctc	cagctggctg	ctccttctgg	aagagccctt	taactctgggt
13681	taatcgccca	tagagcctct	ctctcaatgg	aggaatgtgg	tagctagggt	acagaatgtt
13741	tgtcactcgg	cgctctctct	cttgcttcca	taacttgtat	gggtggggag	ggaagaacgg
13801	tcgtcctctc	tctcttcgaa	tctgtaacgt	aattccactt	acagccagca	tgccccatac
13861	tgccaggata	atgaagtcta	aaaaacagta	gaaaaagtac	agtttgaact	ctagtatttt
13921	aaaaatctaac	ataataatac	aacccttcat	ggactaaatc	tgtatattcc	tttctaatag
13981	actttctggt	gagttagtcc	aaacacatat	gtagtctctc	agttacaaac	caaggaatca
14041	gtagggaagc	cacagtctta	tttttgaaac	taggaatttt	ctattacagg	catttgaaag
14101	tccagacagg	ttgtaaaaaa	atcagttggt	tggatccttg	gccttccaac	attatgtaac
14161	aagtaccac	agttgtttgt	gtgttttttc	agaggggggt	cggggaggaa	ggcagtgatc
14221	tagttacact	tcacatatgt	acatcatgga	acatcagatt	ccaccacgct	catttctttc
14281	ccattccact	ctacatggg	ccctactgca	ctcacagtgt	aggtatagga	atgctgaaa
14341	agattcaacc	aatcgtaact	agcttctagt	aaagcttcta	gcttggaaca	acactgtctg
14401	gtacttaaaa	cttgcaagag	ttttaagggt	ttcaaggatt	tcaaggctc	taccatcaac
14461	ctcttatcca	aagggaaagc	tctggatagt	aaagggaag	atttgcccta	aaataaaatc
14521	acaacaatat	caattgtgct	tatgcaatat	atcttcaatt	ctaggacatg	cttttttact
14581	ctccacatgt	caacatttct	gaagtcaaca	tatcttacia	tctgaacaat	tcattgtgga
14641	gtcatttttt	tcttcaaaaa	atttatagaa	aatctgtaat	ggatagaatc	ttacaagtga
14701	gggaaactat	atgcttacta	attaaaaact	ctttcaacta	gcaaaactga	tgcttcatta
14761	aaggaggga	acaagggaat	aacttaacat	taaatcagggt	tgtcattctc	atcccttatt
14821	ccatccccc	ccataagtct	caaacactct	tactgtcagt	atttttctcc	cctctgagaa
14881	tgaccaggta	aacctaaata	atcttggaag	tttcttaatg	aaattcattt	agtagatttt
14941	tatcttcatc	ttaaaatttt	aaatgcttta	ctcccatgat	cctgagttat	cataacagag
15001	ggttctcaca	gcattgtatg	ctgtgttccc	tccttttgct	ccagttgact	ttcttttaaat
15061	ttagttttcca	agtgtctctg	ttagcatagg	tgggtcaaga	ctcctactgt	gctgtcacac
15121	ttggtttttgg	atttaagagt	cttaccattt	gtttgaaaag	gcacatttgt	gaaagctctg
15181	tggaaatcct	tggtgagcgc	tctcttgagt	acgttcaaa	tgatgtagga	aaggcttggt
15241	gaccagtaac	tgtcaatggc	taaaaccacc	gaataggagc	caatgactcc	acaagtcagt
15301	atgttcagct	gtggaagagt	ggaagagttt	aggcatcatc	catgttagaa	caaaaatagt
15361	ttataactct	ccactggctt	caatcacagg	caatcacatc	tgctccctca	ggcatgaaaa
15421	aggacagac	ccactcagag	ctgaacagaa	gttccctgga	aagataaact	gaaaacaacc
15481	ttcaacatcc	atatgagttc	ttaaaacaca	gttcagctta	tccaaatact	atggaaaaata
15541	tcttccatat	gaaggaactt	caacaataaa	atctaattag	aggaagaatc	actcttctga
15601	agccagttga	aactaatgac	ccatcagggt	ggtgtactag	agtaagcact	accttctctg
15661	ctcccactgt	taaaagcata	ttgactgtga	tgtccattt	gggtgaaaaa	aaagaaaaaa

15721	acaaacacca	tgtggtaate	ataaatagaa	cttcgttttc	acagtaacaa	ggctattttag
15781	aatcagtcac	ctgatacagg	taacacagcc	ctccccagtg	aataaaacat	gcaaaatgat
15841	gaaataagtg	atcagctgaa	cacacaggat	aacataatca	ataaatctaa	cttttgatga
15901	aacacttact	attccttaggc	agcccatgaa	aactactgga	atgaggatag	ctatgcaaga
15961	gaaagtgacc	cagaatacac	catcatcatg	aaaaatcttt	aggtttcctg	aaaaagcaaa
16021	aagaaactaa	gagtaaacia	cagctttacc	aagttagagt	tggtttcttg	gaacatctga
16081	acacagatat	acatcaaagc	cttaaagtaa	caaccaggga	atatgttggc	aactagaaat
16141	gaatttcta	tttcttcatt	taactgtcaa	tacaatcttc	tttgaaggat	gtgttctaga
16201	aggtgggtgg	gaggaaccaa	gcagggtcaac	gctctgcaac	tattaacaaa	atcattttaga
16261	agacactgga	atagtccaag	ctcttccatt	cactagctct	gtgaacttga	gcctcagttt
16321	cctaaccctc	aaaatgacac	tgaattatat	ctccaagctc	taaatcacct	gttagctcta
16381	cattgtctgac	ttcatgattt	catctgtcaa	aataaaaaat	acaagtataa	ggcattcaaa
16441	gttaagccat	gaacccaatg	taaacagtga	gtccacttaa	cagattttta	ccgtcaaagc
16501	atgtcctgtc	atgccactgc	acactgtgac	atgtgatctc	ctcatgtatc	aggaaagagg
16561	aaccaaccaa	ctatctgtct	ttcccaaacc	cttctcctgc	tgctggactc	taaaatgctg
16621	cacaaccctg	gccgaccag	aacaaatata	tgagaactcc	tctttRtaa	ttaagtttaa
16681	aaagcaagaa	taggctgact	gctgtttaat	actgacactg	gcaccacat	tcagttccatc
16741	tgtaaagtgc	cccatacctc	aagacctaca	tatccagggtc	agagtggag	actgggagtc
16801	cctaggctac	aggactggac	cttgagactc	catattctct	gaagccaaca	gcaatatgcc
16861	attgaggtga	acttatagat	tctattctct	agtttaacta	tattaacctt	cacaaaatac
16921	atatgtggct	ttcttgtaaa	aggctttcaa	gaaatctaag	tattgaaaat	agttacgatg
16981	catgcatgtg	aagaacggac	cctttcagag	aataagctca	ttgcttccca	cataagggtga
17041	aaactagcaa	tctaatacaga	ttgtagtgg	cagccaaata	ttttttaaat	gctaattgctt
17101	aaaaatatgg	gtaggcctct	tgccggggca	tggtggctca	cgcctgtaat	cccaacactt
17161	tgggaggccg	aggtgaacag	atcacttgag	gtcaggaggt	tgagaccagc	ctggccaata
17221	tgagaaaacc	ccatctctac	taaaaatata	aaaatgagct	gagcatggta	gtgtgtgcct
17281	ataatcccag	ctactcagga	ggctgaggca	ggaaaatcac	ttgaactcgg	gaggtggagg
17341	ttgcagtgaa	ctgagactgc	gcgactgcac	tctagcctgg	gagacagagt	gagactccgt
17401	ctcaggaaaa	aaataatgtg	tgtgtgtgtg	tgtgtgtgtg	tatgtgtgtg	tgtgtatata
17461	tatataggta	ggccagggtg	agtggctcaa	gcctgtaatc	ccaacacttt	gggaggccga
17521	ggtgggagga	gcctctgagc	ccagtttaag	accaccctag	gcaacagtga	gactctgtct
17581	ctactaaaaa	taaaaRctta	ccaggcatgg	cagcacacac	ctgtagtccc	agctacttgg
17641	gaggctgatg	taagaggatg	gcttgagcct	tagagtgtga	ggttgagctg	agctatgata
17701	gtgtcactgc	agtccagcat	gggtgacaga	aagagatcct	gtctctaaaa	aataaaaaata
17761	tggttttgca	cccactagca	ttgaggagat	gatcactcta	acggctcctg	agctatgaga
17821	tatttcataa	acaaacatat	ataaagggtg	tatgtctaaa	ttatttttac	caatagggaat
17881	agttatgtca	aaaagttag	agacaactgc	tctagactac	tttataaagt	tgaaaattaa
17941	gagacctctt	tatttctgtt	ttctcattcc	tcacatctat	cctatgagat	agtttcacag
18001	atgatggaac	aaagtttcaa	tacatcgctc	aaggttataa	gctagaaagg	gaactagtga
18061	cagagccaga	atgcatatag	aaggtttttt	cctgacttta	taatttatgc	tcatttctct
18121	catgaaaagt	acattaaatt	actagaaatt	ataacctatc	atccctccat	cctctaagca
18181	ttaacaaaag	aggggaagcc	cattaagtag	tagttggttt	ttgtggctct	tttgttttgt
18241	ttttgagaca	gggtcttggt	cagtcgcca	ggctggagtg	caatggcacc	atcatagctc
18301	actgcagcct	caacctccc	ggttcaagca	atcctctagc	ctcagcctcc	cagagttagt
18361	gagaccacag	gtgtgcacta	caacacctgg	ctaatttttt	tttttttttt	ttcagttagg
18421	atggggaatc	actatgttgc	ccaggctggt	ctcaaaactg	agctcaagca	atcctcccac
18481	ttgggtctccc	aaagtgtctg	gattataggc	atgagccacc	atgctcagcc	attaggaagt
18541	atttttattta	gcacatgtag	ggtagagcct	tgaatctgta	tttcaaaaac	atccccatgt
18601	gaatctgtca	tttcgccaag	cctacacaac	actgcttttc	atcagactat	ccagttagtct
18661	tagaaaaaaa	aaagtcttcag	agagacaagg	gacacatgac	ttgtactcta	ttatatcctg
18721	cattctctat	agagttgaca	gcctcatgaa	atggctatct	ttatgtctaa	cagttcccac
18781	aactttattc	tagtatattta	agaacattgt	cacagaaagg	aaacaaccaa	atagaatact
18841	tttgagtaca	gatgctgttt	ttgtttgttt	ttgacaataa	tcaggaaatg	taagttatta
18901	aacaatggca	aagaaagtta	acttataagt	taaccagatt	ttaaaaactg	atagcaattg
18961	aattactttt	taaaaacacc	aaaaaaaccc	cagaaccagc	cctgtgctac	tgccccaggg
19021	tgatttaagta	cctgagggac	acctgggggt	tgccagcgga	tgaggagtga	gggttagggg
19081	acaacagaag	accagaaaca	aagtaccagg	ttccccgact	cagctgccat	cagagccttt
19141	acctccttat	ttcctttacc	tccattatct	ctaataaaca	cataccactt	atcacaacaa
19201	atagatacaa	agctacctct	aaagacattc	tatatctcat	ctactgtgcc	tttgagggaag
19261	gagggacttt	tactagtctt	ggtgctgtct	gaaagtgtca	caccatcaat	gcagattttc
19321	agaactccac	aggagaagca	ctgacatgca	acaccactcc	aggatgcaaa	tccaattagg
19381	cagtattcaa	acagtcactc	gaacacacaa	aacatctgtt	acatatcgcc	taccagtggt
19441	agtaaagaaa	gtcactgacg	agatgaggaa	ccccagcact	agtccaacac	agagcatgca
19501	gatcgagagg	attccaaatc	gccaccacac	agctaccaag	aacattccac	cgacgtttcc

19561	agtgcagact	gtcagaatca	gattcactgt	acaaagagag	aggatgaact	gctggataac
19621	cggcagtagt	tccaaccgcg	tgcccttagac	atacaataacc	gttaccagtg	ccagcttcag
19681	ctttctgtag	gtcaggcact	gagggctctY	cagtgggtgcc	atgaagccag	catagagcaa
19741	ccctctctag	ggaggggtct	gggctgagac	aaggaaatta	gctagggagc	taaacctcaY
19801	agggccagtg	acaggccttg	cttgtcaata	aattgggatt	gagaaggaaa	agttgacaag
19861	cattctgtta	aacataatgg	ggtaggagg	tatctgtttt	caggtaaaga	aaacctcaac
19921	attttgcttg	ttgtacatct	catatcttca	acatatcaca	tagccagaac	acaggctctat
19981	aatcaatatg	gccctcctt	tccttaagta	ccaaggaatt	caccatacag	attaaaaacc
20041	cagaaggtct	gaatgagtag	ttcagaaaag	gtgttcagag	actagatgga	ccattatttt
20101	acataagtaa	taggctgtca	cacacaatat	aggccaaact	gtagaaggaa	agggtgaagta
20161	aaacacagca	cctgccttct	tgaataccca	gttaaaactat	ataattttca	cttaccatca
20221	tacttgatag	gtgtcagctt	tgtaatcagt	atataaaaga	agaatcccat	gatgataaag
20281	cctatgaaga	ataattYtaa	gtggcaaacc	accaggaaaag	gaaaaaagca	aaacaacaaa
20341	agctatggtt	aagcaccctt	tttacagtgc	aattttaaata	tttaatctac	aaatacatgc
20401	ttaagctcct	ggaattttaca	tggtacagca	ttcagaagaa	aaaaaaaatc	taaaaggcag
20461	aagaaattca	gcagccatgg	gagtaattat	tattaaacag	atacaatatt	aaagtaccaa
20521	ctacaggata	cactttcttg	tatataggtc	agtttccatt	gttaaattgt	tatttgttca
20581	tttcatagct	catttttaata	tatttatgtt	ttctgcttaa	aataagaata	gcactttgtc
20641	acgtgtatgt	gatacgaaat	tacacattaa	attttaataag	ttttgtaagc	agaaataaaa
20701	tgtaaaagg	aaaagaaacc	caaactgtcc	caaagcaaaa	tataactttt	atatcttctt
20761	ttaatttcac	ctatcctgtt	tgaatgtatc	aatcattacc	ttcatccatt	aaaacacctg
20821	gtaatgcaag	tatacagata	actttgttcc	cacaaataag	cctaactcctg	atcatggagg
20881	ccatagctct	acctattttc	ttaaattctt	cacagctgtt	cttggttaata	cttgtactgt
20941	atctcaat	caataaatat	aaaatgMtac	atatcaacac	acaaaaaaa	cccagagtga
21001	ggcattggtg	cttgtgcctg	tagtctcagc	tacttgggag	gctgagggtg	gaggatcact
21061	tgaggctagg	aatgtgaggc	tgtagtgcac	tataattgta	cctgtgaaca	gcccatcac
21121	tgactccag	cctaggagac	acagtgcagc	tccatctctt	taaaaagaaa	aaaaaaaagc
21181	tagtgataac	tttgagacca	cagcttagca	actaagttct	ttgttacaga	atcattctac
21241	atgttgagat	attatatatt	ctttacaata	ctctcttccc	aattaataaa	aaaagtgta
21301	taatagcagg	aaaggagtgt	ctattttcaa	caatttaaca	taagttttaa	gagctttatt
21361	agatcaggat	tacatttgag	ctactctgat	gctcagttat	acaacttgtt	gacttaataa
21421	ggatttttta	ataatcgaaa	acatcactta	tataaaaagc	agcccaaata	attcaaacca
21481	attcagtgag	catgattttac	atcccagttc	ttcagtggtg	ttcactgatg	ttgatgtctg
21541	cattcccttt	ggcatctgaa	tctctcagga	ctgagcagac	atttcaaact	cactgtacac
21601	tgagcccca	tagcctgatt	catgtgaaac	caagtcaate	gccatcacca	atatccacac
21661	atatttttac	aaagacaacc	aacttctttt	tttttttct	tgagacagga	tcttactctg
21721	tcatccaggc	tgaggtgcag	tggcacaatc	ttggctcact	gcaacctctg	cctccggggc
21781	tcaagcaatc	ctcccacctc	agcctcccca	tgtagctggg	actacaggca	cacaccatca
21841	agcctggcta	attttttgta	ttttttggag	aaacagggtc	tcaccgagtt	accagggttg
21901	gtctcaaact	cctgagctca	agcagctctg	ctgcctcagc	ctcccaaagt	gctggggatta
21961	cagtggtgag	ccaccacaca	cagccccaat	ttgtctttaa	ttttttggta	gagaagggat
22021	ctcactatgt	tgccctagact	tggtttcaaac	tcctggcctc	acacagtcct	ctcacctcag
22081	cttcccaaag	tgctgggatt	acaggcatga	gccactgagc	ccggccctga	caaccaatgt
22141	ctaattctcg	ttatctttag	tcaactaaca	gatctagcag	ctgataaaat	tactgcata
22201	attgtattcc	cctgtttaata	agagaagggt	atactccgct	gtagggagaa	tgaaaaaaaa
22261	taaaaatgca	ttgcaactca	gtagactcac	gactacaaga	ccccagctac	aaatctatga
22321	agatagtaga	gataagtaag	tgcttactct	cctcacaggt	cattccccag	caacactgaa
22381	aagactgcaa	aatggaaata	gcaactacat	tggacataat	acctgttttc	cagaatctgt
22441	gtccaaagaa	acaaatgaag	aaaccaagca	gggcaaaaag	agtgaagaac	actttggaag
22501	acactcttcc	taaaaaataa	tgaaaattta	gtggataata	catccaattt	gatactttgt
22561	attaatattt	atagaacgaa	ccttcatcaa	aaaagctaag	tgaactgaca	taattttatcc
22621	catacatgta	tattacaaat	ttatctttta	aatactcttg	ccacccttaa	tgaatgaacc
22681	tatgtagaca	atcaatatca	acaactgctg	ctataacaaa	aagagaatcc	agacaatata
22741	tgcttgctga	tagcgggtatg	gaacacttcc	tatgacatat	tcctgggtaa	agtgaggaga
22801	acacagagaga	gagtgttctt	aaaacaataa	atcctgtttc	ttcaacaaat	aaactgcata
22861	gacaaaagg	ggcaggaggt	ggggggagg	tggtgggggt	gggtggagg	agaatacagg
22921	ttaaaagaga	tttaaaagac	acatcaacta	aatactgtgt	atagatctca	tctgcaacct
22981	gattcaaaca	aactgtggaa	aacaattgag	atgactccag	aagcaaaact	tggctgttaa
23041	tttgatgcta	ttaaagaaat	attttcaata	ttttaaaatg	aaaattatgt	taatgacacg
23101	gtggttctat	tttttaagtgt	cttatcttta	aggtatacat	actgaaatat	ttgtgagtaa
23161	catgatgacc	aggattttgc	tttaaaaata	tccagtgata	aaggagggtg	atgaggtgga
23221	gaggcagaga	agcagggtaca	gcacagctga	aacaagataa	tggccataaa	tcaaaattgt
23281	taaggctgga	taattggtgt	atggaggctc	gctgaagttt	ttcacaataa	aaaattttaa
23341	ccgaaactca	caatttaaaa	acatgtatta	cacttataag	ttatttatta	taatttttat

23401	tacacataca	caaattacta	ccagtcttat	caattatata	tattgccttg	gagaaaaatg
23461	ttgcattaca	accaatgatg	aagcacggta	tcatttagtt	tactgaagag	aaaaatatgc
23521	tcactaaatt	catcacccac	catcatgaga	taaaccacaag	aaaataatcc	ctgaactaaa
23581	acccaaaaca	agacatctat	acaatctcat	gctaaaaata	aacaaggctg	ccacaaatca
23641	aaatacataa	aatgcctaag	agtgtctcaga	atggagaaac	catgaagaca	tacagtaaaa
23701	taagaaacac	agacagaaga	gcagaattca	caggggatata	aatacctaaa	tatcagcttt
23761	tctgagaata	aggagcccat	aagaaaagatt	aaacaacagc	cgggtgcagt	ggctcacgcc
23821	tataatccca	gcactttggg	aggccgaggt	gggcggatca	cttgagggtca	ggagtttgag
23881	atgagcctgg	ccaacatggt	gagaaccccc	tctactaaaa	atacaaaaaat	tagccaggcc
23941	tggtggcaca	tgctgtagtt	cccagctact	ccagaggctg	aggcaggaga	attgcttgaa
24001	ccaggaagc	agaggttgca	gtgagccaag	gatcaggcca	ttgcactcca	gcctggatgt
24061	cgcagccaga	ctccatctca	aaaacaaaaa	acgaaaaaca	aaaaaaaaaa	accagaaaag
24121	attaaacgaa	agataaaaaac	accacattgt	ttatttcaga	cctgtggttg	actatcattc
24181	tagaattttaa	tttttacacc	aaaaccatta	aatttgctgg	gatgattttc	tcaaatgaaa
24241	atcccgtgaa	cttttcagca	agtataatca	ccaagacagc	aaggctctcc	aacccaataa
24301	aaactaagac	acagttagat	actcttccat	aacattagag	gggaactaaa	tcccctatca
24361	caataccagg	aaggaggag	agaaaacaatt	atacaagtaa	aatgagaggg	tatggaaggt
24421	tccatcaaaa	gcagaatcta	cgataaaaaca	agacataaaa	caggcattca	ctctggggga
24481	atgaaaggag	tgaatagact	aggacaggag	caagttgaag	gttaaagtgt	ctttaactat
24541	gacagtattg	ctaagggtta	tgaaccacag	ggcaaggttt	tagttaggag	gcaatatgga
24601	tatagtctcc	ctactccaag	ctcaggaaaa	gcactctaata	actcatcaat	cttgcaaatg
24661	agagtaacaa	tagaactaac	atcactagga	aggttcaact	ttgggcggga	agaacaactg
24721	gatcagagct	ctctgtctcc	taagcctaac	agagagagac	aggcctgggg	gcctttctat
24781	ttagaagggg	cttttgcatg	aaggggttcaa	gggaaagaaa	tggaggaaacc	cagacatact
24841	tcagtccatc	ggttaaaaat	atatggccag	gcgcagtggc	tcacgcctgt	aatcccagca
24901	ctttggggag	ctgaggcggg	ttgatcactt	gaggccagga	gttcaagacc	agcctggcca
24961	atatgggtgaa	accccatctc	aactaaaaat	acaaaaatta	gtcaggcaag	gtggcgacaca
25021	cctgtaggac	tttgggaggc	cgaggcgggc	atatcacctg	aggtcaggag	ttcaagacca
25081	acctggccaa	catggttggt	cagggtctcta	ctaaaaatac	aaaaattagc	agggtgtggt
25141	ggtacacgcc	tataatccca	gataactctg	aggctgagggt	acgtgaatta	cttgaacctg
25201	agaggcagtg	gttgccagtg	gtgagatcga	tgccactgca	ctccagcctg	ggtgacaaga
25261	gtgagactgt	caaaaaaaaaa	aaaaaaaaaaaa	aaaaaaaaaaaa	aaatatatat	atatatatat
25321	atatatatatac	acacacacaa	acatatctgc	caaggtctct	tgcatgtgat	ccctcaagga
25381	tcttgagtgg	aagattgtct	tgagtggaaag	actgagtgggt	gacaaagagc	ctttgcttgg
25441	ccaaatttct	ttttcgtttt	tttttttttt	tttttttttt	tgagacagag	tctcgctctg
25501	ccgcccaggc	tagagtgcag	tgccgcgcatc	ttggctcact	gccagctccg	cctcccgggt
25561	tcacaccatt	ctcctgcctc	tgccctcagcc	tccttagtag	ctgggactac	aggcgcccac
25621	caccatgccc	ggctaattgtt	ttgtattgtt	ttttttttta	gtagagacgg	ggtttctactg
25681	tgtttagccag	gatggtctcg	atctcctgac	ctcgtgatcc	gcccgccttg	gcctcccaaa
25741	tggtctgggt	tacaggcgtg	agccaccgca	cccgccctgc	ttggccaaat	ttcagtctcc
25801	tgaaccccct	cctaggccca	tctgcgcaca	tctttgtaaa	atccagcttc	agcaagaact
25861	ctgataagtc	agcttagcaa	gaatccccta	tgcttgatat	caaatacccc	tcgatatcta
25921	atcgagttct	tcatcctcca	ccatccatcc	cccaggtgat	gtctgaccac	cctgggtctgt
25981	cttcagtaag	aatcccctta	ggtcagtgca	gacagaatgc	ccccttgccc	ctcatgtttc
26041	ctcttagaaa	ttttccatcc	actgactccc	accctactgc	ttggctacaa	attcccagta
26101	gcccagtttg	tactcaagag	ttggggcccaa	tctctcttcc	ccactgcaaa	attccatcac
26161	agtggtccct	atgcctattg	taatactacc	ccccaccttg	aataaaatct	tcctcactgt
26221	gctttaacta	gtgtcactga	tttttttttt	ctttaacagc	gggatacggc	tctagcagag
26281	gtagagaacc	ataatctctc	gcttgaacac	aactaataga	ctttcttctt	tagagcaaaag
26341	ctccttgaaa	gagctacttg	aactttctcc	agtttctgtg	ctcccagctc	ctatttggtc
26401	ttgaatatgc	tccagtcagg	cgccctacccc	acctctccac	tgcaacagct	ctcactggta
26461	ccaatggcct	ccatgttgct	aaatccaagg	gtcattttca	gtctctttct	tagctgacct
26521	atcagcagca	tctgccaaaag	ttgattactc	tctcttccat	gcaatactgt	ctccccttgg
26581	cttcaacata	catcactttc	ctttttaccaa	tctggctgca	gcttctcagc	ccattggtag
26641	gactttcaat	ttcctaagct	ctaaacattg	gagcaccagc	ggctcagtc	tcagaccctc
26701	cccttttctc	taagtcacac	tgaggtgtgc	tgccctctcc	acctcccctt	ccacctctgc
26761	ctccccttta	tagataacac	catttagcat	gaggccttaa	atgccatata	ttcagctctga
26821	agattttcaa	gatgctctcc	agctaagact	ttcctctgaa	ttccaggctt	atatatatata
26881	tcaaccatca	gttaaacatc	tgcacataaa	tgctcaatag	ggttctcaaa	ctgaacttgt
26941	ccaaaactga	actcttaatg	tatcctccca	accttctcct	ccaacagtgc	ttcccacctc
27001	agcgaatggt	cgctctctct	ctccagttgc	tcaggccaaa	aactacaaaag	tcctttcac
27061	ctcttttctc	caaaccctc	atccaatcca	tcagtacact	gtactccaaa	catatccagc
27121	cactcatgag	ccaaatctgg	aacaatttga	acatgaaaat	aaatgatagt	aaaggatttc
27181	atccagagaa	taaaatagga	atctacaggt	ccaggtttaa	aataaatata	tggataaata

27241	aatggcagg	gaggggaaa	ccttcctcag	agcaggatgg	caacaaataa	acatagaagg
27301	agtgtacat	tagagaaaca	ccatttggca	accacctagt	aataaatgtt	tcaggcaaag
27361	atcactggat	gccaaaacta	gcaagtgaga	gtttgagaat	gggatattta	ctgacactcg
27421	aagtacctcc	cctaaagaca	cttaaaaaaa	ggaacagtta	cttcagtggg	gaaacttggc
27481	aagacaccaa	taaccaagtc	atcaaggatg	tcaccaccag	taaagagaca	aataaataaa
27541	gagacaagt	tctcctgaga	aggacacatc	acttcctttg	tattcctgcc	aaaaaagcat
27601	aatcaaaatt	taattatgag	gaaatatcaa	agtaacccaa	attgaagaca	ttctaaaaaa
27661	aataaatatt	ctgtacttct	caaaaaagtc	aaggtcttaa	aagattaaaa	ttaaaaaaaa
27721	aaaaaaagac	ttgttccaga	ttatacaact	ataaagagac	acatcggtta	aatgcaatgt
27781	gggatcctaa	actgaaccct	ggacagggga	aaaaaaaatt	tgttgacggg	gagacggggg
27841	ctcactctgt	tgcccaggct	ggagtgcaga	agcacaatca	cggctcactg	cagcctcaac
27901	cttcctgggc	tcagggtgat	ttcccacctt	agactaccaa	gtagctggaa	cgacaggcat
27961	gtgccaccgc	accagctaa	tttttgtatt	ttttgtagag	acggagtttc	gccatgttgc
28021	ccaggctggt	cttgaactct	ggagctcaag	caatccgccc	atcttgacct	cccaatgtgc
28081	tgggattaca	ggcgtgagac	actgcaccca	gccaaaaata	atTTTTTTTT	acttataaag
28141	gacactattg	gaacaactag	taaattgtgt	aacgtgtgaa	ttaaataact	actgtatcag
28201	tgtagcttct	gtaattttga	taatttttact	atggttatgt	aaggaaataa	atacagaagt
28261	atttaaatgg	aaaaaggtct	ttatgtctgc	aacttactct	catatgggtc	agagaagaaa
28321	attgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgta	cgtatctata	catacacagt
28381	cgggccctga	acaacacggg	tttgaactgc	atgggtccac	ttacatacag	actttttttc
28441	agtaaaaagt	acactgagtg	tgctgcctcc	tcctacctcc	ccttccacct	ctggcccccc
28501	tgagacagca	acaacaaccc	ctcctcttcc	tcctcctggt	cctattcaat	gtgaagacga
28561	tgaagatgaa	ggcctttatg	atgatccact	ctcacttaat	gaatagtaaa	tatactttct
28621	cttccttatg	atttccttaa	tgacatttac	ttttctctag	ctcattttat	tgtaagacta
28681	tagtctatga	taaatataac	atacaaaata	tgtgttaatc	aagtatttat	gtaatcaata
28741	aggcttctct	ggtcaacact	aggctaactg	tagttaattt	tggggaagtc	aaagttatac
28801	acgggtgggg	tatggtggct	tatgcctgta	atcccagcac	tttgggaggc	caaggcgggt
28861	ggatcacttg	aggccaggag	tttgagacca	gcctggccaa	catggtgaaa	ccccgactct
28921	actaaaaata	caaaaattag	ccagggatgg	tgggtgtgtg	ttgttaatcc	cagctactca
28981	ggaggctgag	gcaggagaa	cacttgaacc	tgggaggcag	agggtgcagt	gggccaagat
29041	gagccctact	cactccagat	tgggcaacag	gagtgaact	atgtctcaga	aaaaaagaaa
29101	agagttatac	acagggtttc	aactgcacag	gagtcggctc	ccccacaacc	cctacactgt
29161	tcaaaggtca	attgcatata	gagagagaag	aagagagaat	gaaaagcaaa	tagcctaata
29221	cattaagctt	tgaggaatct	ggtatgaagt	tatacagaaa	ttcttttatac	tattttttgca
29281	tattttctga	gaaaccattt	caagataaac	tgctaaatac	acacacacac	acacacaccc
29341	ctcacaacac	cttcactgcc	ccatttctac	ctaaccctag	tatctctccc	ctggattact
29401	tcagtggcct	cctaaccaat	ctgcatttct	gctcttgcc	ccgtcttttc	tcaaaaagca
29461	gccagtgatc	tcattaaaa	atcttagatc	atggcaccat	ttgctcaaaa	ccttccaatg
29521	cctttccctt	tcattcagaa	tagaagcaaa	atcccatact	acagtctaca	aggccacaca
29581	caggctgcct	gcccctgcac	ctctcacctc	acctctactt	cctttgcccc	cctcgttttc
29641	ctgcaacccc	tgctatcact	gcagttcttc	acaccctcag	acaggctccc	agggtcaggc
29701	cctcatcctg	gctgctccct	ctgcctggta	catgccacac	ccaggctacg	gctgtgtgca
29761	tcaaggactg	ccctggccat	gctatctatc	tcaaattttc	ctcttactgg	catgtgttat
29821	ctccctttcc	tacttttatt	atcttcttag	cactcaccat	tttctcacgt	attacatatt
29881	ttataaaatt	atcttgtctg	tctcatctcc	ctcactataa	tataagctcc	agaacaagta
29941	tttttgcgtg	tttgtttact	gctatgtccc	agtgcctaga	acagagagtg	gaataacgta
30001	ggggctcaaa	aggatgtgtg	tgaatgaatg	agaactagac	aaaagcatta	tctgcaaggc
30061	caaaggagaa	tttgtaaaa	tttaaagcca	catactcacc	tagggaagca	caactaccct
30121	ctcctgcctc	aaagctgcaa	gcgtatgtgt	gagcaggaat	gtaggcagca	gatgtattta
30181	gaaacgggtc	ccaaacaatg	acattgtata	tgacaccttg	tcccgggagg	gaggagaagg
30241	aaacacttgt	cttatcatta	gctgttaggg	taaccacctg	aagccaaaaa	taacacatga
30301	ttagtgttta	ggggattttt	ttaagctatt	ctgtgataat	cccacctacg	tgtaatatata
30361	aatggtatat	tcataaatta	taattcttgc	ttaaaatata	caccaaccca	tttaatatgg
30421	tcaaaaaggcc	atcttacatg	gtttgtatgt	gctaattgtaa	acaagctctc	acacaaaaatg
30481	cttccaaaagg	tttttttttaa	tagttttaagg	aaaatgcatg	ataaacaatg	tataaatgct
30541	agttacgacc	atgaaaaaag	ttttttcacc	tctgctagac	tataaacaag	gatgggaact
30601	gtatttctct	tgtgcaactgc	tctttcctca	gtgtacagcc	caatgcctaa	cacagaggat
30661	gtgtccaatg	ttagccacag	actaacgcag	acagtctggt	aacatctcgc	agacagtacc
30721	cacttatttt	tgtctaacag	tctctgaaac	aagaaaatat	tacatgaggt	tttgttatgt
30781	agtagtagtt	ctttttttttt	ttttttttttt	tttgagacgg	agtctcgctc	tgctgcccag
30841	gcgggactgc	ggactgcagt	ggcgcaatct	cggctcactg	caagctccgc	ttccccgggt
30901	cacgccattc	tcctgcctca	gcctccccag	tagctgggac	tacaggcgcc	cgccaccgcg
30961	ccgggctaatt	tttttgtatt	tttagtagag	acgggggttc	accttgttag	ccaggatggt
31021	ctcgatctcc	tgacctcatg	atccacccgc	ctcggcctcc	caaagtgcgtg	ggattacagg

31081	cgtgagccac	cgcgcccggc	cagtagttct	taacttcaga	aattccattg	taaactataa
31141	aatagaagaa	acaaatgctt	tatagaacca	taatacctat	gttcctacta	atttgagaaa
31201	aataaatatt	agggttttag	gagaaaatta	ttatgagccc	tcctacatca	gactaccatt
31261	gtttaaaaat	actgcttagc	cttcaataac	ttgcttcctg	ggcaaaggct	tggtaaagtt
31321	caacaaactg	ttaatatatt	tcaccaagta	caatgaaaag	tttaaaccag	acagaaaagt
31381	tgaatgaata	acacaataaa	tagactatat	actgcctaata	atcaacaatt	gttaatatatt
31441	tacatttgct	tctctatcac	tagtttctcc	ccaactctac	ttttttcctg	ctgaaagtaa
31501	gttacagata	tcatgacatt	tcacttctaa	acacttcaact	gtgcatgtcc	taccacagt
31561	tttctcaaat	ttaacgtgct	tgtgaatcgc	agggggcagt	cttggttaaaa	ggtagatctg
31621	attcttaggt	cagcgtggta	ctagcaattg	tgcctttcta	gaaagctctc	agatgatgcc
31681	aatgtatcgc	aaacacctac	cacactctga	gtaacaagtt	actcagtaca	gagtaagtc
31741	taagtacaga	cttacaagg	ggttacatca	ggataaacc	atcgtaaagt	caaaatgcat
31801	ttaatgtacc	tcacccacca	aacattatgg	ctttacccta	gcctaccta	aatgtgctca
31861	gaacacttac	attagattca	gttgaaacaa	atcatctaac	acaaaggcta	ttttattttt
31921	ttattattat	tattattatt	attgagacgg	aatgtcgcct	tgttgcccag	gctggagtg
31981	agtggcacga	tcatctcggc	tcactgcaag	ctctgcctcc	caggttcacg	ccattctcct
32041	gcctcagcct	cctgagtagc	tgggactaca	ggcgcccggc	accaccacta	cgcccagcta
32101	attttttgta	tttttagtag	agacagggtt	tcaccctgtt	agccaggatg	ggctcgatct
32161	cctgaccttg	tgatcctccc	gccttggcct	cccaaagtgc	tgggattaca	ggcgtgagcc
32221	accatgtccg	gccacaaaat	gctattttat	aataaagtgt	tttaattgctt	atgtaattta
32281	ttgaatcag	tacactgtaa	aatgttagct	attaccctgt	tgggttaagca	ggagcgtagg
32341	ctcactgcc	ctgctcagcg	ttgcaagagt	atcatactgt	atattgctag	ctcaggaaat
32401	tatcaaaact	aaaatcaaaa	gatcataagt	caaaccatca	taagtcaggg	accatctgta
32461	taagaatatt	ctcctatata	acaatacctt	agctcaccta	cgaaaatgaa	taaYttctta
32521	ttagctaata	aatgatccag	attttcccta	atgtcttttt	tttttttttt	tttttttgag
32581	acagggctct	actctgtcgc	ccaggctaca	gtgcagtggc	gtgatctcag	cttactgcaa
32641	cttctcccag	gctcaagcga	tcctcccacc	tcagcctcct	caatatctgg	gactacaggc
32701	acgcaccacc	acgcctggaa	ctttttgggt	tttttagtgg	ctgtgaggtt	ttgccacgtt
32761	gccaggctg	gtctcaaaact	cctgagctca	agtgatctgc	cctccttagc	ctcccaaagt
32821	gctgggatta	caggtgtgtg	ccactgttcc	tggctcctaa	catcttttct	agcttttctc
32881	cttcttacc	aagatcaaac	caagtcccat	gcatggtgtt	tacggccaat	ctcagtaatc
32941	cagagcagtc	tgtcccacct	ttctctccca	gctacccaat	gactgtagaa	gatgtgaagt
33001	cagtattctc	ataaagtatt	ccacaatttt	gtgtctaact	tattctttta	cccccatat
33061	ttcctgtaaa	tagggaatcc	aatggactaa	attcaggtta	gacatttttt	gctagaatat
33121	accacagggt	atgggtgata	gtgcagactg	cactacatRa	gcacgtaata	caggctgatc
33181	catttttcagt	gaaactagct	ttgccacct	ggaatagggt	gagactataa	ggctactcca
33241	tgttaaagtg	gttctccttt	tacaatgaga	attaatctaa	ggatgctact	tagacaccct
33301	agaaatatcc	tgttccccaa	aaacctttca	cctgctgatc	cttgctggaa	ttatttcatt
33361	gagggtcaca	aaatgggtgat	catctaattc	tattatttct	tctacatata	ttagccagca
33421	ctcttctgca	aagagctttc	tctcaccaac	tggggaagaa	ttagtctctc	ctttaaaaag
33481	gcaagctaaa	tttctttatc	aatttttaga	gtaatttgct	gtaagagtct	cctccagtgg
33541	tggcaactga	ggaagggtttc	atttgtattt	attcaaggct	gacaaccaca	gtctttttct
33601	tgattcttga	actgccccaa	atttggcagt	gcaagccctt	tcaaacaggc	tcctgaatct
33661	ctttgacata	ccctcattag	tctttgagg	tttcccaca	ttctggtaca	aaatattctt
33721	ttaaactaaa	gtttaaaagt	aacctataaa	aagcctatta	tacaagtctt	tatattcttc
33781	ataaaaaata	taatagtatt	aaaatgtcat	atctaataaa	gcaactaata	aactttacta
33841	acagagcttc	ttattttgtg	aataagcaca	atgacgctgt	ctagaaaaac	tcacccatc
33901	aaccattaga	ctgaaggatg	acctcatata	ttatattgct	gttaggcaat	tttcaagaca
33961	taaagataaa	atcctgccc	ccacaaaaat	gatctattgc	agtgagtctc	aaaatttaat
34021	atgcataaaa	atcacctgag	ggctggagt	caacagcacc	atctgggctc	attgcaacct
34081	ccacctcctg	ggttcaaaaca	attctcctgc	ctcagcctcc	caagtagctg	gtattacagg
34141	catctaccaa	catgcctggc	taatttttgt	atttttagta	gagacacgg	ttcaccatgt
34201	tggacaggct	ggtctcaaac	tctgacctc	aggtgatcca	cccgcctcgg	cctcccaaag
34261	tgctgggatt	acaggcgtga	gccactgcac	ccagttagaa	tcacctgagg	gatataaaat
34321	gcgaattctg	attcagtagt	tccagggtgg	agtttaagaa	ttacaattct	cacaaactcc
34381	ccagtgatgc	caatgccact	ggtccatgg	ccacacctgg	ggagcaagg	tttacattga
34441	ttcaaaagct	cagagaagaa	ttttactcca	agttagactt	tgagtttggg	ttcaaatatg
34501	aatgccaatg	tgtatatcct	taagcagttt	ctcattattt	ttttctcaaa	acaatgtaat
34561	cttcatactt	gccaaaaatg	tttaccag	agtcgtggga	gtacacagaa	aattaccacc
34621	tcctttccatt	cacatttcag	attaccaagt	cacagatgtg	gctgcaacta	aaacgtaaca
34681	taattttacat	gcatctcctc	aaaggccaac	atcattttga	gatatacaca	gcaaagactc
34741	ccaagatgat	aacaaaagag	aataaagtga	gtgtccccc	aaaaggagaa	aagagtaaag
34801	aaataagatc	ctggtagctt	ctccccttct	caactctgac	aaagggtgca	caatccagcc
34861	acacaaatca	tgggactcag	atttaccttg	agagcactgg	ccttcacctg	gggcacactg

34921	accatcctct	gcagatgctt	cagcaacatc	tcctcagtga	ggtcattctc	aggcagaaaa
34981	tactgataga	catcatactg	caacctccac	ctggagtcct	ggctctgtccc	agcgtcacat
35041	gggtgggggat	ctacgcctct	aaagtcaacc	aacaaaaatga	aacatcagtc	tcttgtgggtg
35101	caagtcccat	ctataaaaaag	gaaccaagag	ccctaatttc	aaacacttgc	aacaataatt
35161	acaatttaac	agttcacctt	ttgcagtgcc	aacaattggt	attctgaaga	aaagtaatgc
35221	actgttatgg	actgaatggt	tgtgtctccc	caaaaattaat	atgctgagac	cctcactccc
35281	agtgtgatga	tattaggagg	tgggaccttt	ggaaggtaat	taggtcatga	agggtgaagcc
35341	ctaatagaatc	gaattactgc	cctcataaag	gagatcccag	agagccctct	agccttcttt
35401	ctgtcatgtg	tggacacaat	gacaacccag	cagtctgtaa	cctggacgag	aaccttcacc
35461	agaactcaac	tgtgtctggca	tcctgatcgt	ggacttctac	cctctagaac	catgagaaag
35521	aagtttataa	gccaccaggt	ctgtgctact	ttgttacagc	accacaaact	aagacaaaac
35581	ccaacaggca	aagggtgatta	cataactgga	ttctgatata	agtcaggcct	agagtaaaat
35641	actacagctc	tggggaggca	aaaaaatttt	ttttaaaaaa	aagctgggtg	cagtggctta
35701	tgccctgtaat	cccagcactt	tgggaggctg	aagcaggagg	attgcttgag	ctaagaagtt
35761	caagaccagc	cagggcaaca	tagaaagacc	tcactcttga	tttaaaaaaa	aaaaaaagaa
35821	gaagaatata	gcagatttgt	catttgttag	caagtagggt	aacctccata	acctctaggg
35881	aaactttctg	taaaataaca	gcaattgtccc	ttcttcataa	aattgttagg	attaaatggc
35941	aaagaacata	acaacactca	gcaccacgga	caatgcacag	gaagcattca	agtgttctca
36001	aatgtttagct	atcattacta	ttaaagcaga	aactgaggca	cagctttaat	taaaacaatg
36061	aatctttata	aaaataaaaag	gatgtttcag	ggctactgag	tgctatctat	atttaaagtg
36121	ctttataaaa	accaggatgt	attttcaagc	tcataataga	ctcacatctg	gctcttcaaa
36181	taaaaagagt	tcacaagaat	gaagtgggaa	aataatgaat	gtaaatgaca	gtccaaatat
36241	taatagttaa	gtttaaatta	catcaattcc	ccccaaaaat	taactcaatt	ttactacaga
36301	gaattaggga	cttgagagag	ctggtcaaat	ctacacaact	gtcgtcatgt	atattctaaa
36361	agcatttctc	agacatggga	gacagacgta	cctcgcatag	cctaggtttg	ctggggcaaa
36421	cttgatagtc	gtttcaaaga	aattatactc	caagtaaatg	ttgggatcaa	tatctaaatc
36481	gaactccaaa	ttacagcctc	cagggacagg	atctggataa	gaaataggga	agttactaaa
36541	caaacaattc	ctacatatcc	aaatcatact	cagaaaaaca	gaactctatc	tactcttaca
36601	aaaacttccc	aaatagaagt	taaatataaa	cgtagagctt	ctagtcgtat	actatttcat
36661	gaactgaggt	taaagagggt	agagttaact	gaggttcagg	gccctagagc	agggcagagc
36721	caaaattttc	tgtaaaaggc	ctaatagtat	tttaggcttt	gtcagtcctg	aggtctctgc
36781	agtaactacc	caccccagtc	cttgtggcag	gaaagcagcc	acaggcaata	cgtacaggaa
36841	gaggcatagc	tgtgtttcag	taaaacttta	tgtacaaaaa	caggcagcaa	gctggatctg
36901	gtcctaggtc	tgtagtgttc	caacctttcc	tctagagtta	aaaagatcgg	ctattgggtg
36961	actacttctg	gccagggttg	gaatgcagat	acacagtgtg	tatatatgtt	tcaatatcat
37021	gttggaaatc	tcctccaaat	tctatgctct	tccttcatac	tctccttttt	ctgcacttat
37081	ggatctttac	cacaccacaa	atggtatgta	tgccagctga	agctcctctg	ttcataattt
37141	ttctttgtac	actgtccttg	acgcaagcct	taattaagtc	tatggtttct	tatcaggtaa
37201	ttagcctttt	gttatagcta	acatttccta	tcttctactc	cagtaatctt	ctggattctc
37261	cctcctttcc	taactagcgc	tcctctctct	tccttttacc	cccaagcttt	aagagtatgc
37321	taaattttatt	cacctcacct	tcactctctc	atccctgttg	tgccactaaa	gtgactctca
37381	taaagggcca	ccgctgactc	ctacttgctc	tagcatacgg	acactttcag	tcagttctct
37441	tgacctctcc	gtagactgta	acttaccact	tccttttttt	tttttttttt	tttgagacgg
37501	agtcttgctc	tatcacccaa	gctggagtga	agtggcgcaa	tctcagctca	ctgcaacctc
37561	tgccaccg	gttcaagcaa	ttctcctgcc	gcagcctccY	gagtacctgg	gattgtaggc
37621	acctgccacc	acacctgact	aatttttgta	tttttagtag	agatagagtt	tcgccatggt
37681	ggccaggcta	gtctcaaaact	cccgattata	taaattatac	ctgcaacctg	cttcggcctc
37741	ccaaagtgtc	gggattacag	gtgtgagcca	ccgtgcccg	cctgctccct	tttaaaggca
37801	gaaactccag	gctttctctc	taccttcttg	actactccac	agtgttggtg	ttattcctca
37861	ggagttatta	tccaccagaa	acagtgcctg	acacatagta	ggcactagta	aagtatttgc
37921	tatgattttt	ttgcctaaa	tatccaagta	catgcctgga	taatctcact	cactgccaca
37981	gcttttatta	acagctatct	agttgggaat	ctaactctac	ccctccactc	caggccctta
38041	tttaaattat	gtactctgca	tttccatcca	aatgtcctac	aagcatctca	gactgtgtcc
38101	caaactgaag	tcataccta	ctccgtccca	cctccaaaaa	catccagcac	caccataacc
38161	ctgttctctac	acctaatttg	tatcactatt	tcctgtctgc	ccaaatagga	aacctatgct
38221	tcctctttat	actocataac	caccagtcct	tgtgttgtag	caattccacc	ttctggatat
38281	ctcaaactct	gtcctctctc	tccatcccg	ctctcatggc	ctctgttcag	atcctcattt
38341	ctcattttgt	gctgagtcct	gactctgctg	cgccctctcc	aatccatccc	cacactactt
38401	ctagagttag	ctttctaaga	gatgaaatgg	actacaccat	tcctctgctt	aaaaactcta
38461	tgccctccct	gaacctacct	cgaataaag	tacaaacctc	aaattgtatg	caaggccctc
38521	cacactccag	ctgcagccac	actcacaacg	tgccattctc	caaactcatt	ttcttcatga
38581	cgctcagctt	actgcttctg	cttagaacia	ctaaactggt	aaggactatg	tcttgccctta
38641	actattactg	cttctccaa	ggctccctaa	atgtttttcg	aatgaataaa	catgttctcc
38701	tagtcactac	ctgagctttc	agtcactcgt	ctaattctct	tactataaaa	tatttctgat

38761	tataggctgg	gcgcagtgcc	tcattgctgc	aatcccagca	ctttggggagg	ctgaggcggg
38821	cgaatcacaa	gggtcaggagt	tcgagaccag	cctggccaac	atgggtgaaac	gccatctcta
38881	ctaaaaatac	aaaaaattag	ctgggctctg	tggcgggcgc	ctgtaatccc	agctacttgg
38941	gaggctgagg	caggagaatc	gcttgaaccc	gggaggcaga	ggttgcagtg	agccgagatc
39001	gcgccactac	actccagcct	gggtgacagt	gcgagactcc	atctcaaaaa	aaaaaaatcc
39061	gattatataa	attatacttg	caacaaacaa	tctcaatagg	ataattttcc	ttaaagtttc
39121	atgaacacat	tccatattcc	aaagagtaag	agggaggaga	agacagaagg	gttacctctt
39181	tctgagtagg	agagtaggat	agccatattc	tggacaggct	gtatgcctga	agtccccaag
39241	taccaagtgc	atgtactctg	ctctgggtcta	aggatgaaaa	ccagtccact	ggcagtgccct
39301	gtttccgagg	aattggaaaag	gagagtctgc	agtccagaag	aaacattgtc	atgaatagga
39361	actcgacca	aacataaatt	actgagctga	tccagtgtgc	atgtcctatt	ccaaatcttc
39421	ccaaaacaaa	tagaaaaaaa	aaaccttcct	taaataacca	gggaattctt	tccatctcta
39481	tgttcacatg	cccacacata	tacatggaaa	aagcattctt	gaaaaatata	cgattgtctt
39541	agttaaggct	gctataacaa	agtaccacag	gtaaggtggt	tataaacac	agaaatttat
39601	tttcgcaatt	ctagagctctg	gtgggttcag	atagcgggtg	caacacagtc	gggttctgat
39661	gaagagctgc	ttataggctg	aagactccca	acttcttgtt	gtatcctcac	atgacagaga
39721	aagggaaga	aacatatctg	gggtcccttt	tataaggcca	ctaattccat	tcatgagggc
39781	tctaaatggg	ggcccatgac	ctaatacact	cccaaagact	ccacttccta	atataacaca
39841	ctgaggatga	gtatttcaac	tgtgaatttt	gtgaagacac	aaacattctg	tctgtaaaaa
39901	taatctttga	tcaggaatag	cagatgcaca	gtgaaattcc	caagtgtttc	aatttaattg
39961	aattaaagatt	tactatgtca	aataatgttt	tatcttagtt	tctaagacca	ctaggacaaa
40021	aacatgagac	ctggccccaa	ggagttcagt	tttcgaggta	ccttttaaat	tgtgttataa
40081	ctgacaaact	ataaactgta	cataaagtat	aatttttctt	tttttttgag	atggagtctc
40141	gctctgtcac	ccaggctaga	gtgcagtggt	gcatctcag	cttactacaa	gctccgcctc
40201	ccgggttcac	gccattctcc	tgtctcagcc	tcccaagtag	ctgggactac	aggcgccctg
40261	caccacacct	ggctaatttt	ttgtactttt	agtagagacg	aggtttcacc	atgttagcca
40321	ggatgggtctt	gatctcctga	cctcgtgate	cgccagcatc	agcctcccaa	agtgtctggga
40381	ttacaggcgt	gagccaccgt	gcccggccca	taaagtatac	ttttaagtgt	tgatatacag
40441	acacctgtga	aaccattacc	acaatcaaaa	tagcaaatgt	atgcatcacc	ccaagtttcc
40501	ttgtgcccct	tcctaatect	tcctcacaca	cttctctctt	accatcccc	actccatctt
40561	gcttggaag	ttttttgaat	aataaaatgc	taagctaaaa	ggtaacagac	tcaagaggtt
40621	ctatcaaagc	taaggaaatg	ggaaggaaaa	ggtagcagag	tagtttcaat	ttgataagca
40681	taagaaagaa	tgactcaaat	attaattgat	aatgatcaa	gtcaacagtt	tcaaagagcc
40741	aattgccatc	acgactcaaa	aaaaacaaaa	gtgccacact	ttattaggaa	gggcacagga
40801	aacagaagca	cttctcttta	cacagaccac	tcacttcctg	agccatgtgc	agttctggcc
40861	atcatcttct	agtaaatata	acagtgggaa	ctaaaggcag	tccaagaaat	aaaatgataa
40921	ttaaggacaa	ctaagaccct	ttctgcatag	aatcaaggc	ccaagatgct	gccagctagg
40981	tctgttttat	cactgggtgat	gctgacatct	ctttatccaa	gttctgcttc	ctcagatgta
41041	gaaataggca	tcactactaa	ttgtttgagc	tttaatgaga	attaaccata	tggaaaccaa
41101	acacaggggt	tcctaactgg	gggcctgaga	gtctgagttt	ttgaagggcc	ccatgaaatt
41161	ttatatattat	atttgagttc	atagacttga	gcagtctatg	gtccaggag	cttaagaacc
41221	actggcctag	gacatagcag	gcgttcaaca	aatataaaYc	cttatttcac	ccatcctctc
41281	tcttttcctt	caaatgttgt	cttgctgtaa	tgtcatttag	aagaaacaYa	ctgtttgtca
41341	attaaaaatt	gttttaatat	atctagttaa	aaaagagaag	aagaaagaac	tcattctatgt
41401	gccatttttac	taagtctcct	atgtaagagt	ctgaatagga	agctgtcttt	tgaattcagt
41461	gccataaaaa	acttggcctc	ccttgaggaa	ctggcatgct	ccttcctttc	taagggaagt
41521	atgaaattta	acatgcttta	ccttctgctt	tagctgcaaa	gaacttgggc	aaattcgacg
41581	tatgtgtgac	atcgggtatct	cttgtaatta	atactattta	ttcctttcct	tttttcctaa
41641	gctgaattct	tccattttcta	gtctattaac	taccccccta	cacacatacc	ccacaacaca
41701	cacgcacaca	cgtgcacaa	cacatgcaca	caatttccaa	acctctatgg	taggggaaag
41761	aaggcaggaa	taaaaatata	caaaaatagta	agaaagatag	tgtgaccaat	gcctctaata
41821	ttatatatgg	gctggggcgc	agtggctcat	gcctgcaatc	ccagcacttt	gggatgccca
41881	gacaggcgga	tcacttgagg	ccaggagtct	aagaccagcc	tgggcaacag	agcaaaaccc
41941	catctctaca	aaaaaaaaat	gttaaaaaatt	agccaagtgt	ggtgacatgt	gcctgtagtt
42001	ccaggatcac	ttaagaccag	gaattttgagc	catgatcaca	cgaccacatt	ccggcgtggg
42061	tgtaggagtg	agatcttatc	tctataaaaa	aaaataaaaa	ttttaaaaat	ttcatgatga
42121	attagattaa	gacagaattg	tttaccaaa	tccagaatat	tataatatag	gcttaggaca
42181	aataaaaaaga	agtataactt	catacaagg	gaaatgcatg	tattggtcta	gggcatcata
42241	gagactaaat	gcagatagat	tgtaaaagat	tataaaaatg	gcagactctt	catagcttaa
42301	actaaccag	agaattttta	aaacttttcta	aggctgaggt	tgagggttaa	gggttttagtg
42361	atgacatcta	taataagagtc	cttaatactg	ctggaggcat	agcacagggc	tggatgaact
42421	gagcactata	atTTTTTTgt	tgTTTTgcgtg	acagagtcca	aagaataaag	ttgatattat
42481	agggttcttg	gtaaagttca	atactatagc	taaaacaaca	tgaaaaccca	ttaaacaatat
42541	attctcaggt	aaaagttatt	aaggcccccac	taaagtaaga	aaatgaaaag	gtgggaattc

42601	acgaagaaaa	cttgccctgca	tagtgggtgac	atltgtcaca	gggtgcccctt	ctccccagct
42661	cagctctcca	ttacccccttt	ttaggtcaga	gactgccaca	ctcaagaaaag	gcctgagata
42721	ctctcagtc	ctaacttgaa	aacatccctat	cttgattttt	taaaatttta	gacataagat
42781	cataaacata	aaagagatag	agatatcaaa	ttctaattgct	cataactaaat	accagacccc
42841	caaaatatag	aaatgatttt	tactaaaata	attcttttgca	cttaccggag	aaaaggaaaac
42901	agttgtattc	tgatactgtg	agtgtatttg	gaaaataaga	aaagtcacat	tgcttgaaat
42961	atcatgcaaa	atagcttcct	ctggaaaagg	cctattgagc	tcgaagtatc	taaatttccc
43021	cacagaaaat	tcaataagac	ctacaacgag	ataaactttt	cattacaaca	cttgactctt
43081	catatcaatt	tgagaaaatc	atattcataa	atgaatgttg	aattttaatcc	tcagaaaaac
43141	tccttgaggt	aaatatgggt	attgctccta	tttactgct	aaagacaatg	tggttaaaaa
43201	agtgatgcag	cccattccagg	gtcatccagc	cagtgaatgg	tagagctcag	aactaaatg
43261	gtctaaggat	ttcaaaaagg	atgctgaaga	caggctcctg	gaacccagag	aaacttcccc
43321	tctggttagca	gcaataacaa	tgatctctgg	aacagagtcc	ctgaatttta	agatgacaat
43381	cctgtggggc	tactaacaat	catgaatgtg	ctggtgttaa	agaaccataa	ccgtctagtg
43441	aattctgaag	tcagcgtagc	ccgaagagac	acaggtggct	gcaattttaat	accggtgttt
43501	accagctgtg	tgatttttagg	caggttcctt	aaatcccctt	aagcctcaat	ttcctcatct
43561	gtctaaggga	gactgcaaag	gaagtgcaaa	tcaataactga	caaaaaatgtc	caatgacac
43621	ccacaccagg	cactgcacat	aaaattgtga	accacaggca	ttctggaaaa	gagagctttt
43681	gtttgcttaa	ggatcaaaact	gagtcataa	aaaagatagg	ttataacaag	aacagttttt
43741	caacacctcc	tccaggcctg	tcccactctg	ctacttgcc	ccatggctctg	catctcactg
43801	actctgtccc	cacttctctg	tgatgtgttt	ctgccagctg	cacaccagct	catctttttt
43861	caatctaact	ttcaaaataat	cccttaatta	ggttgctttt	actcctgtgc	ggctcctggat
43921	tctaccacca	cccccatcc	tacagtcagt	atctgtctgc	atgaacagca	gaaaagatcc
43981	caagtttcac	ataagcagtc	ttgggggttc	aattctaaac	aaacagacaa	aagatttaga
44041	ttaaatatag	atgaaaagga	attgatgaaa	aaagattata	attcaccgga	gtcccatagt
44101	ggaactgtca	tgttttattt	tagaaggtaa	ttagagtcag	tgggcggtaa	aacatttcag
44161	actactaata	acttaataac	acattcttat	ataccatgtc	ccaaaaagg	caattttcaa
44221	cacttttttt	tttaatacga	aagctccact	ggatgattta	aaaccaaaca	gaaccaaaaa
44281	aatcaagaat	gagtagcctg	tggaagaaca	gccaaaaaaa	agacacaaat	agtgttaaga
44341	ctagaataat	cttaacctct	atataatttt	cctaagctct	gtgatgctac	attacattca
44401	taacttaatt	tttttttaaa	aaaatgatag	aataaaaact	tacaattttca	ttctaaaaac
44461	tctagagaaa	attgtaatat	ttttaacaat	gacaaatgaa	gatagcatat	ttcctattta
44521	tggtaccaga	aagacatttt	acatttttaa	tggaacaaac	tccagctgga	acagagggtta
44581	aaatatatta	gaatggccaa	tagaatctgg	gtagggaagt	ctggaaaaag	tttgtgtgca
44641	cgtgtttttt	cttttttaaaa	agaaacaaga	aaaacatgtt	tatattttcc	aggccttccc
44701	attatcatga	ttattttatt	agaaaagcaa	attatacata	aaactatatg	gatttccagta
44761	aattacacta	aaaatgtttc	attttattgc	cttcaagttt	gggtcccatcc	atcatgcac
44821	ctccaagtct	taaaaattcc	caccgcgtatt	ggcagtatca	gtctgctttc	ctagaaatta
44881	atcagatttc	tatcctttca	ctgttcacag	cattcaactc	acttactatc	ctgccttttc
44941	tagtctacat	cttcagtttt	tcttttttgt	tctcattttc	cctattttggc	aatattttata
45001	aggtagctcat	ggactaggca	ctctagtgtg	tgggcaccat	gatcaaaaact	gacacaaaaca
45061	aattcataaa	ggttatccaa	ggcttctgta	attatcactt	aaatttgccc	tttgatatta
45121	ttttagaaaa	tggtcttaata	atcctcatgt	ctatctccca	tgacatgagc	cttccatcag
45181	aacagtatac	agaaccttta	aaagccctgg	aagagactga	agccatcaac	tagagaagcc
45241	ttcatggtaa	actttttctta	acettccctt	aacctatcac	ccaggatgag	gatgaatcac
45301	aactcaaaaa	tacttgagaga	aacaaaactt	tggaagagaca	gtacggaaac	acagaacccat
45361	ggaaactaaa	atacattggt	aaacaactga	tgacttgaca	aaataagtag	gggggtatta
45421	aaagcaggta	tctcttgag	tatattttaa	ctgatttttt	tttttttttg	agatggagtc
45481	tcactctgac	acctagggtg	agtgcagtg	cgcaatctta	gctcactgaa	atttccacct
45541	ccagggtcaa	gcaattctcc	tgctcaacc	tcctgagtac	ctgggattac	aggcatgtgc
45601	caccacgcct	ggctaatttt	tgtattttta	gtagagacgg	ggtttcaacta	tggtggccag
45661	gctagtctcg	aactcctgac	ctcaagcgat	ccgcccactct	cagcctccca	aagtgtctgg
45721	attacaagca	tgagctaccc	tgcccgctg	atltttcttt	tttttact	tgctctctt
45781	tttctttttt	tttttttttt	tttttttgag	acagagtctt	gctctgtcgc	caggctggag
45841	tgcagtggcg	caatcttagc	tcactgaaat	ctccacctcc	agggttcaagc	aattctcctg
45901	cctcaacctc	ctgagtacct	ggagtctacg	gcagtgtcca	ccagctcgg	ctaatttttt
45961	gtattttttg	tagggactgg	gtttcactat	gttgccagg	ctggtctcaa	tctcttgacc
46021	tcgtgatccg	cccacctggg	cctcccaaag	tggtgggatt	acaagcatga	gccaccacgc
46081	caggccaact	ctacatttta	aatgtttata	aaataggaaac	acacactaga	aaaaaaaagg
46141	gacaaaatgg	gacattctct	caacattctg	tcactgtttt	gctttccctt	aaaatggtaa
46201	tatttttcta	gaaaatatat	aatccttata	aaagtgctta	aataaacaga	aagtcaactc
46261	ctccagcaag	tacaatgtac	agaaatcacc	actgctaaca	gttgggatat	atcttttagg
46321	gcttaggcag	atataagcat	gcatatttcc	gcatagatgg	aaaaagatat	aatatgtttt
46381	gatattgccc	atttaacagg	tcttatatat	ctttccatag	caacacagat	aatgctagct

46441	ttaccattttt	aaaagctaca	gagtattcca	ttctgggtat	gtaccataat	tattcaccca
46501	atctcctatt	aatgaacatc	taggctgtta	ccattatagc	atgaataata	aattaataag
46561	aaataaaaata	gttctgattt	tgtggaagac	tattccagaa	agcgtatgtg	tcaggcaata
46621	aatctattttg	agagagtttt	tgctgtgaca	gtgtaacaga	tgacctttga	ccttaaattg
46681	ttaaactgta	cctaaaaagt	cttgttctag	tctaagatat	ccagaaataa	aaacaaatag
46741	accatgtgag	aatacacaat	gattagggac	aaaaattaat	ctgcagtcct	tgcttctgaa
46801	ttcttcttca	agaaatatac	ctgaatttca	ggaataatta	caactttcat	ctgttatttc
46861	ctggattttcc	caaaagctac	aggaaaatga	tgccctcatca	tcttcctttc	ctcaaagtcc
46921	tacagtgcga	accaggcaga	agccaggaaa	taaaaggag	ggacactgta	tacattatgt
46981	aaaacctggt	cacttgacaa	aataggtagg	ggtattaaga	ggaagtatct	ctgggggtgt
47041	attttaagt	attttttttc	tttccattac	acatttcttt	atttcttaaa	tttctatagt
47101	gaccataagg	tgctccttttc	ataagaaatt	ataatacagt	gattatacac	cctggctatc
47161	cattacaatc	atctggagg	gattaaaaaa	aaaacagAAC	actgctgcct	gggaccctct
47221	ccagatgagt	taaatcaaaa	tccctggggc	taaggTccag	gtatcaatta	aaaaacaaat
47281	atccattaaa	tactatgcag	agaaatacag	aaagtttaca	cacccatgaa	ttaaccagca
47341	cctaAACagc	attctagcat	cttagtggtc	cccctgtggc	ccctcccaat	cacatcttct
47401	ttcctcccca	gagggaataa	gggtcctaag	ttcaaacact	aggaatttat	ttggacgaat
47461	tcaattgaat	agaatttaca	gcacaatttt	tgtgtgtttg	gcttctttaa	aatcagcatt
47521	atgtttgtgc	aattcattcc	agctgctata	ggtagctata	actcattacc	ataattgtat
47581	acttcatgaa	ataacatgcc	acaattatct	gttctactgt	taataaacat	ctgtatttct
47641	ggctttagagt	tataaataac	tctaAACatt	ttctgtgcat	gtcttttggT	acacgcataT
47701	acacattttt	gttgggtgga	acctgtgagt	agaattatca	ggtaatagg	tattctcatc
47761	ttcagtagat	actgccagtc	ttccaacaat	gttttgcaa	tgtataccgc	atcaacttac
47821	tgtaggaaag	tggtccaatt	gcccagaagt	ctcaccaaca	cctggtatca	ttcatttatc
47881	tgctgagtat	atagtgggat	ctcattgtga	ttttaattcg	cacttccta	attactcatg
47941	ggtgttaaga	aacctttgct	ttgcttattt	gctatttgga	tatcctcttt	gtgaaacagt
48001	ctgttcgagt	cttttacaca	tattaaaaat	ggaactgtcc	ttcttcaatt	ttaggagttt
48061	ttaatacatc	atagatatga	gtattctgac	agatacgttg	cgaatatatt	ctcccactct
48121	tgttgcttgt	cttttcactc	tgttaccttt	agattaacag	ttttcatttt	agcagtcaat
48181	ttcagggtttt	tcctttatga	ttagtccctt	ttgtgtcatt	ttaaagaaat	ctttgtcaac
48241	atataccaag	atcatgaaga	tacattccca	tgttAACttc	cagaagtttg	tttctctttt
48301	cacattttaga	gctaaaaatc	aagctgatgc	tgatttttac	acatgggtga	agatagaggt
48361	cacagttcat	tgtttcatat	gttcatctac	ttgacagcaa	ccttcactaa	aaataccacc
48421	ctgtagtacc	ccatgttctc	tcacttataa	atgggagcta	tgtaatgtgt	acacatgggc
48481	ctacactgtg	gactgataga	cattgaagac	tgaggaaagt	gggagggggc	ctgaagatga
48541	gaaattactt	aatgagtaca	atgtacataa	ttccagtgat	ggatacacta	aaagccctga
48601	cttcaactgca	caatatatcc	tgtacctaca	cttgtgtctc	ataagtttat	acaaataaaa
48661	taattttaaaa	ataccaccat	ctccccacc	tactgctttg	cagttccacc	tttcacataa
48721	gtcaagtgaa	catatgtgtt	tgcttctaga	gtctcaattg	cattctactg	gcttatttgt
48781	ctagctatag	acaccaatcc	taacacaaat	attacaactc	tcttaactcc	tgctccaaat
48841	ttgttcttcc	tcctcaaat	tatcttgact	atttctgata	ctttacattt	ccatataagt
48901	tttagaaatc	agcctgtcaa	tatccacaca	cacacaaaaa	aacctgtctg	gatttttgatt
48961	gggatggccc	tgaatatata	cattaattta	ggaagaaatg	atacctttaa	aatatcttta
49021	tctttagaat	tttgtaatct	atgaatttcc	gatattcctc	tatttcttta	gtttctcaat
49081	ttatctaaag	aatgttgaga	tgtattatgt	tattttctct	gtagaggtct	ttacatatcc
49141	catgattatc	atcccagaat	acgtaatgtt	tcttgaggct	atttttaaag	tacttttctc
49201	tttctttaat	tttatttctt	aacgttttgt	agagatgagg	tcttgctata	ttgatcagac
49261	tggtcttgaa	ctcttgggcc	caagagatcc	tccccactc	agtcttccaa	agtgtggga
49321	ttacagacgg	aaccacagca	cccgaccctg	ctttttcttt	taaatgtcta	tttctagctg
49381	tctcttcacc	aatatatctt	ttataaaaaa	ttctccatgt	aattctaagt	tttagctggg
49441	tttgagaact	cctaacatag	tgtttgataa	acaagcaaa	gataaatatg	tatttttttc
49501	ccagtcttag	ttatctctaa	cctaacatct	gtgtgaaaa	actttggcag	tgtaacctg
49561	gcaatctaac	atccattgta	caattttttt	cagaagattg	ctttaaccaa	accagttgtg
49621	gtttgattcc	tggtgtgtga	ggattttgtc	atttttgatg	ttccaacgtc	taacacagta
49681	tagtcaacta	actggcatac	agtaaatgtt	tgtaaataga	tgccctgaact	atgtgacctc
49741	aggcaagtaa	tgtacttctc	ctgggcttca	gttccctcaa	agtaaaattg	aggtaactgt
49801	gctgccatca	gttggtgtgt	gaaagtatca	aacaaatggg	aaacaacttt	taaactgcc
49861	aaggctacac	aataaagcat	catttattaa	tttccattag	ctttcctgtg	cttcacacat
49921	tcatctcctt	attgaaagaa	aagaaagaaa	gaaagacaat	caggcactca	gagcagccct
49981	acgaaaaaat	tagtggaag	aaaggaacta	acatgtacag	ggcattcaca	gaaattccag
50041	gcacatacata	agcattacgY	tgagaaagag	acttccgaag	agtgacaaca	aaagataactg
50101	aaagtcaaga	caacttaaac	agctcagttc	ggtagaaact	catttaattt	ctgctcataa
50161	aaacagattt	agcaaaagaa	gacacatact	caaaaaatat	cctgaagaag	ctaacaggaa
50221	tttaacttca	cctctaaggc	atatggaaaa	ggatttatta	gaagcatttt	atgaggcata

50281	attatatctt	atgacttgca	gagtagggag	aaaaaagaca	gcacatcaaa	agatattaca
50341	aagcaaatgg	ctaacatcaa	gttgggggcac	ttacattagc	tttcctctaa	aaagaaatgg
50401	ctttttttgc	catctgacct	atagtaggga	cttgaaacaa	tttttcacac	ccatcttcca
50461	ttcttaccga	gaatcacttt	ttctttccaa	taattccact	ttaaaaagag	ctgccccatc
50521	ctaataacttt	ccttttttctt	tccgaatcag	atcctaactc	ggcaaccata	aaaatttcca
50581	actcttcctt	tttactatatt	tactaaaatt	tacttctact	gtttgtgttt	caacacaaca
50641	accaaattct	aagagaaatc	catcatagcc	tagtcttcat	aattctcttc	acattaaaaa
50701	caaaaatggg	tcagctactt	gagttgggca	gcagtcaccc	gcccctccca	atttctcatt
50761	gttatcttct	tacaaatggt	caatagccct	aaaggttgaa	agggaaacaca	aaaatgaaca
50821	tattagcttc	acacactttt	tagagaaaga	cagcacttgc	aactaaccat	gaacagcctg
50881	ctaccttggt	aaagcatcat	caatgcgaag	aagactgaaa	ggcccaaatg	atgctgtgct
50941	caagtccaat	tactatgaaa	ttccacggga	gttactggat	atgaatctaa	acaaatgaaa
51001	gatgtttaaag	gcagcagtga	actgggactc	caagagaagg	gaaaactgct	taaaagtatg
51061	atccatctca	tgtgcctact	ggtcacagaa	agcaccctcc	tcagacattc	taggaactcc
51121	atctgccacg	ttccaagatt	tatttagcat	tcaggccact	acagaagtca	tgcaagttag
51181	ctcatcctgc	caaagatggg	taaatcctcc	taaattacag	cagatgtcta	atttcttcag
51241	tttgtatggt	tatcagaatt	atcatcatca	aacatcaatc	cagccatttt	ctaggaggat
51301	tatgcaagtc	ctgaactatg	tccttctacc	tttctatact	gatttttagc	tgatagatgt
51361	ggagaatatc	agctttttatt	atcactcagc	agtttaagtcc	ccacattgtc	cctcttctgc
51421	acagaatgac	aataccaatg	caaggcaaaa	agaataacat	tgaggaaat	cagtgtcata
51481	atcatagcaa	agaggcaaaa	agcactgtct	gaaaaaaatt	aatttctcaa	tttattcata
51541	ttttttaact	tttgtataaa	aatttctagg	tattaaatac	taacaccaag	tagaaagaaa
51601	aagttctgga	cgctgtctct	cctgactcag	gggctggcct	cgatcaacta	aaaactctaa
51661	tatctcaaat	tcttcacctg	ttaaaataga	gatagcatct	cttctacctc	catcataagg
51721	ctattttgaa	gaatgaggat	tcacagaagg	attaagtaca	aagcaccatg	ccatgtaaga
51781	tattcttatt	ggaagacagc	tttacggaa	tggaaggact	atgaaatttg	caggagact
51841	cacctgggtt	caagtcctat	ctgtaaccct	agcagctggt	catttgttcc	tctgacaagg
51901	gaggaagatc	attacattgt	agggtgggtg	ggattaaata	acactgcacg	tggaaagcac
51961	ctggcacaca	ctatacagag	tgaatgttag	tgccctatta	taagcaaggt	ttttctgcaa
52021	aaaaaaaaagt	tgcccttttc	agctttggcc	tcagtccctat	atacaaagat	cctagaattc
52081	agttttgcaca	agggttttcc	aaagataagt	gcatagcac	acaatcacat	tattcagaaa
52141	atattatggt	tcagaactgc	aactagtcta	ttaacctcaa	aacaaaacag	aatgtgctta
52201	gtttcatgga	aaagcacatg	atgatctgcc	tggcacaact	agtgcctttt	acattttata
52261	ttaacaaacc	atgagctggt	gccccaaacc	attcaccaag	cacaagactt	acaatctaac
52321	atccccattc	gagttctcac	tgagatgaac	aagaattacc	tgaaatcggc	cgggcgcggg
52381	ggctcatgcc	tgtaatccca	gcactctggg	aggccaaggc	gggcagatca	cgaggtcagg
52441	agttcgagac	cagcctgaac	aacatggtga	aaccccgctt	gtactaaaaa	tacaaaaact
52501	agctgggtgt	ggtggcgggc	gcctgtaatc	ccagctactc	tggaggctga	ggcaggagaa
52561	ttgcttgaac	cgggtgtcgc	gaggttgtag	tcagccaaga	tcatgccact	gcactccagc
52621	ctaggcgaca	gagcgagact	ccgtcaaaaa	aaaaaaaaaa	aaaagtcacc	tgaatcatc
52681	aaaaatgggt	tactcatagc	gagccaagcc	cgggcattgt	gcttcttggc	tcccacaggt
52741	gaatttgaag	tggcaataga	acttcgttga	ttcttgtttt	tctcatttta	gttttactct
52801	ctattaaatt	gttatttttt	acagttcatt	tatggatgac	aagttatttc	ctgttggtag
52861	ctgcaatgat	tgggaatttt	gtgatctctg	taaacgaatt	aatatctcat	agagatctat
52921	ttcctgttgg	atatgagaca	acagagaaca	cagtttttgt	ccatttgtct	tttttaaggt
52981	caaatcgatt	agaaatttct	tgccctcaag	taaggagaac	aactctgata	cctggagact
53041	caagcttttg	tgtttctgtg	tttatttttc	atttgtggct	gaacttgtag	aattgagggt
53101	atacacattc	tatctgtatg	tctatgtatc	tgtaattcag	aaagtccttt	atttctcatt
53161	gtgagtatgg	aatgttttcc	tgtctccaga	tggtattaat	aaattagatg	gtaaaggctc
53221	tttaaaaaata	aagcctcatt	atgggttcat	ttccctctc	aaactctgta	gaacataacc
53281	agtttgaaaa	gYctctagat	ttgggggtgt	aaaaattcca	tgcagcttga	gagccatcag
53341	tcattaagcc	gctagtcca	ccccaaagtc	tgctgtccag	aagctcagac	agYccactgg
53401	gcagcaaaga	ggcgcttttag	taagttgaat	taaagccaca	aacatggctg	ctccctaagg
53461	gagataatac	tgtgttaaaa	aaagagaatt	atttatctct	tccattttcc	tccgctaaat
53521	gaggagggtta	atagtaagac	ctccttagcc	atgagtgttt	ccgctgcagg	aaacaccatt
53581	cgtgccgggt	ctggactcac	cacccagcgt	gtaccaacgc	ccccagcacc	atctccctgc
53641	taggctgaca	gacccctggc	ggattttgac	ctcgcaagaa	atgcaaaaagc	aactttgggt
53701	tgcagaagcc	aggcgccccg	accttaccct	cgctggaatt	cccgaagacc	tgggctgcac
53761	cagccaccgg	gtgttcggat	gccagcaccg	ctacgaccag	cagctgcagg	aaccccatg
53821	ctgcctgggc	cggtctgggc	ccacgcaccg	gctggggaga	ggtgcggggc	tgcgcggccg
53881	ggccccgcag	cctgcgccac	gctatcccg	ggcgcccgca	tggggcgcca	tgcggcgcca
53941	ggtgcagacg	cttcgcacct	gccagctccg	cagccgcggg	cgcgcgcccc	gccgaactcc
54001	tagccccagc	gagaggtttc	ctcttcgggc	acctgacccc	caacgcgcgt	tcccgcgcct
54061	cctcccgggc	tctcgctgcg	tcacagcggc	gggtctctgg	gcgcggggcg	cgctggggcg

54121	tttccaggca	gcgcgcgcct	cctgccggcg	cgcggggaac	gcgcagcctg	cggtgctccc
54181	gggagccccc	aggaagtcct	ggcttgaggc	ccggtcttgc	cccactttta	tctaccaag
54241	taggcaaacc	cactgtgaaa	gtacaaaata	actttctgtt	gcggactaac	gtcgccaggc
54301	tcctgtttcc	gtgtttaacg	gccctttaca	acagcaaaaa	agaggaaatca	gttgctaaaa
54361	acacaattga	agggtatttc	ccccggtagc	aattgtcttc	gctgctgagg	cttatgttag
54421	cataaaaaatg	ccaggagaat	gctgtttgtg	tagtggttac	cgcactagtc	ataagcacta
54481	catttaaaag	tagtttcgtt	ttctctaagc	gaaatatcct	cgaagacacg	gccttggtct
54541	gaacttcgtc	tcaagtctca	tacctgagtg	ctgtttgagg	gacagtagtc	catgtgacag
54601	atggctcatc	gatgatgtgg	aattagacgc	atagacttaa	cgttatttgg	ggaggattgt
54661	cttgctttcc	ttccctgggt	ttaatttttt	tcacttatat	attttgatga	gcaacttata
54721	gcttggaat	tacatgttca	gttagtccac	tactaagcta	acaatatata	gctgtgacca
54781	taatgtgaag	aagtgttaaa	gacaaaatta	ttcaatgata	cttggttaaag	caccgaacga
54841	agactattga	ggactgtctc	aattggcata	gggaccactg	caactgggtc	ttgcagtggg
54901	gaagaggggac	tgggctcaac	tccgaataca	gcgtgggcaa	gagggaaattt	atagccaacc
54961	agcagtatgg	agttcagtg	atggaaaatt	actaagagga	agcatcagag	gtaaggggca
55021	tttggtgaaa	ctgacctaac	aggactctcg	aagacataac	atcacctggg	ggatgggtga
55081	gggtgagaaa	actgattaga	tatggaggaa	agggggttct	ttgctaaaaac	tggattttta
55141	caaagaagtg	cacagatgga	cctagcagaa	gattcagaag	ccgtactaga	gtttggccaa
55201	gcaaataatc	tttgtcatag	gcacaattct	ttctctgctt	tcccactgcc	tgggtccaaat
55261	tttcacacct	ttggtagcac	tatctatctc	tgtagtgtct	attcactctc	ccacatgccc
55321	ctattcctat	aacctcagt	agatgaaat	caagggaagat	gaaagggaca	attatgataa
55381	gggtagcatc	tgaatcagtg	gatggttat	tcaataaatt	gtgttcttaa	ctgattagct
55441	gcactatggg	cagagagaag	cagaactcta	tctcacacta	attgtgaaaa	caaaatccca
55501	aagtaataaaa	accataaaaag	agctagcaga	agagggcgct	cttttttttt	tttaataaat
55561	tgaatcctaaa	atgggaggag	ctttctaaga	agcaaaacag	caaaggaaac	aaccataaag
55621	gaaagagattg	gtagacttgg	ctactgtaaat	gctttttaca	taaatatatta	aattttatta
55681	aagtaaacac	acaaacaata	tgaaaaaggca	agttgacca	aaaaaaaaaa	ttagcaaagg
55741	acatgaatgg	ggaatttcac	cccaaaaaag	aaatacaaat	taccagtaaa	tacatgaaaa
55801	aagtccaatc	tcactagtaa	acaaataaaa	aaatcaaatt	aacaaactaa	gatgtaaaca
55861	ttttaaaagt	ataaaaacatc	cagtgtggat	aaaagagcta	agaaggaggc	attcttaggt
55921	agggatactg	ggaatataca	cttaataaaa	gtaatcgcca	tcaaaaatgt	ttaaaaatct
55981	tatttttghta	gatttgggtat	aatccacttt	tagatttaag	tacataatta	aatacacata
56041	taaaaatttag	tataagctgg	gggaagtgg	atgcgcctgt	agtcccagat	actcaggggg
56101	ctgaggcagg	aggattactt	gagcccagga	gttcaagagg	gcagtgtgt	atgattgcac
56161	ctgtgaatag	ccactgccct	ccagcctggg	caacatagtg	agaccccatc	tcttaaaaaa
56221	aaatagtgtg	cagaactatt	catagaaat	tatttagagt	agcagaaaac	tctaaaacaa
56281	cctaaatgcc	caacaattag	gagttaaata	acctatgata	gagctttata	agaggatcct
56341	gtgcagccct	tcatttatat	tctgattgaa	gtctttatga	aaaggcaagt	tatgatgcaa
56401	taagtatata	tagtacRtgg	cttttttttt	ttttNaacat	acatctttta	ttgactcaga
56461	ggaattggata	cttttgttgt	ttgtttgttt	ttttaaatgg	agtcttgttc	tgtccccag
56521	gctagagtgc	agtggtgcaa	tcttggtcca	ctgcaacctc	tgcctcccg	gttcaaggcaa
56581	ttctcctgcc	tccgcctccc	aagttagctgg	gattacaggc	acacaccacc	atgccgggct
56641	aatttttghta	ttttttttta	gtagagatgg	ggtttcacca	cgttggccag	gctgggttgg
56701	aaatcctgac	cttgtgatct	gccacctcg	gcctcctgaa	gtgctgggat	tacaggcgtg
56761	agccaccaca	cccgccctgt	aatttttttt	taagagacaa	gatctcacta	tgttaccag
56821	gctggcctta	aactcccaga	tgtaaagaatt	aaagaagaaa	gaaacaggaa	aggtggctta
56881	gaagtcaagg	acaagtttat	tttagagaaa	acaaacctga	aaggggcttc	tggccagtta
56941	ggtttagaggc	acactctctt	acggactaat	aagagttttt	aaggatttag	ggtgggagag
57001	tttatcagag	gcttggactg	cttctgtgtc	tgttgtgtct	atctgggagg	gagagtgtg
57061	tgtctgttcc	catacatctt	tttgcagctg	caggcatacc	cccagctctg	cttttagctt
57121	ccctatctta	gtgcacctga	agggaaagga	atgtacttac	taaggccac	tgttttactg
57181	gggcccattg	tatgaggggtg	aagtttggca	gttaccacaag	agaatttccc	cccacctccc
57241	tttgtgcctg	agctgtctta	tctgtgtttt	actgtctgct	ctttctggct	gcttatagtt
57301	agaagagaag	tgatttcctt	gaaatgcatg	aggctagaaa	gggagctgga	acttaaagt
57361	gcgggtgtttg	tccgagatga	cggtgtctct	ggtctgtcac	cagactcaag	caatcctg
57421	acctcagcct	ccagactaac	tgggactact	acaggctcac	accactgtgc	ctgcttgtt
57481	ttctttgtag	tatgcattcc	tggttatgta	catagaaaaa	aaggttgaaa	aatatatgcc
57541	aaaatgttta	aaatagttat	ctttagggtg	ttaaataaag	aatgatttcc	tttgacatt
57601	tctgtgtaaa	agttctacat	tgaagtgtat	catttttaaa	taagaaatta	tgattttgta
57661	aaagatacat	ataattttta	aggttttctt	ataaaaagcat	tcaagatata	agtccacctt
57721	cctgagtatg	acacataaat	gctttttgtg	atgcaactcc	tgctcacctc	cctaggttct
57781	tgctgcaact	aaaatcagac	ctctacatgc	tggccacatt	gaattaattt	caatttctctg
57841	aactgagcct	attctctctc	ctccatacct	ttgcatcgca	gttccccagg	cctgcagttg
57901	cctgtgcatg	catgacagtt	gatgacaaa	gtgtcagcaa	tcagaacagg	gcctcattac

```

57961 cttggcttct aagtacagta ttgtgacagc ttgggaggat gattggctgg ggttccatgc
58021 tcatgtcaga gttccattgg agagctgaga agcaaaacag cttagcagct ttgcagatta
58081 gttcattttt cctatgtatt gaaaggaaat gcagaaactg aaagattccc aaatccaaga
58141 tatgcatcat tagaggctta aagtctgttt aggccaggcg cagtggctca tacctgtaat
58201 cccagtactt tagcagacca aggcaggagg atctctggag accaggagtt caagatcagc
58261 ccaggcaaca tagtgagatc cccattctta cacacacaca tacacacaca cacacacaca
58321 accagaaaaa aaaattagct gggcatgggtg gggcacatct gtagtcctag atacacaggt
58381 ggctgaggct ggagaattgt tctgagcctag gagtttgagt ttgcagcgag ccatgattat
58441 gccactgtac tccagcctgg gtgacagagc aagaccctgt atctaataaa acaaaagtct
58501 gttaaaattt agaagaggaa ggaaggcgag gagggagaga aggaggaagg gaaaaaactt
58561 accctagctt tttaaataaa acctaaaata ccttaaatgg agctgggagt ggggatccag
58621 ggaaaaggca ggattgaacc tctgtgaaat ttgcccattc attctttccc actctaattc
58681 cacccttcac ctgtattttt taattcttat gctcttttgt gggtagtaat ttacattctt
58741 tgttcattaa gttaaattca gtaaatgctt gtttagtagc tattatatgg aaattattat
58801 gctaggaaac tcattcattt ggttgataaa tatgtgccag gctaaaWgtt aggtgctagg
58861 tactggaaat atggaacag aatatgtttt ctgcctgcaa gtagcccatg gactgggtgg
58921 tgaattagcg tgatgtacaa ataattacag cacagtatgc taagtgcaat aacagaagtg
58981 taaggaacaa ttgtgttaaa gattgggagc atttaacttt cagaattcag gaaaggcttt
59041 tataaagaga tgtttaaaat aagattttgt tccactagtat agtggcaagt aaaaaaata
59101 ataattaggc gggcacagtg gctgacgcct gtaattctag cactttggga ggccaaggcg
59161 ggtggatcac ctgaggtcac gagtctcaaga ccagcctggc caacatagca aaaccccgctc
59221 tctacacaaa tacaaaaaac ttaccagggt gtggtggcgg gcacctgtaa tcccagctac
59281 tctggaggct gaggtaggag aatcactaga accctggggg cagaggttgc agtgagctga
59341 gatcgtgcc a ctgcattcca gcctgggtaa cagagtgaga ttctgtctca aataataaaa
59401 taaataataa aataaggttt gagtactgtg tggacaatca ccaagggagc aaagtggaga
59461 tgtggatgga agtgtaggca tttcatacaa ggaagaaata attatagtg ttttatatg
59521 gtgtgcattt ttgtaaaaca gcagaacatg tttagggaat ttcaaagagt gtggtataga
59581 aaaatgctta tttYagtta ttcattccgg cactaagaac agtttagata tttataaat
59641 tgcaactgtg agtattaaag Ktatttaaaa attgctaaag gtgtttttct taaaaaatag
59701 ttgtctacag cataaaaaac aaagcattta caagattgac tgagtggctc actgctgaag
59761 ctgtggaggg tctgcgaggg ttcactatac ttactacttt tctatatctt ttttataatg
59821 aataatataa agtttcaaaag aaagtactca ccaaaattgt gtgtttaaaa tatgagatct
59881 gtaatactgg gttcctgctc ctaaaaccat ataatccagt tggcgatgca acagaagcaa
59941 atggaaaagc atatgagaaa gtgtccaaat catgaaaagg agaaaaggaga ggtccctatg
60001 gactctagca attcataaaa ttatatgag aagtgtacag cttaaaaagg tagggtaagg
60061 attaaatagc aggaaagaaa ttgctttggg aaaggtaaag aatgaggaaa gagctagacc
60121 agtggttctc aattgggggc aattttgccc cccagcaaca tttgggacta tctggagata
60181 ctttatgttg gcataaatgg attgggggga tgctaccggc atctcagggg taagggacag
60241 ggatgctcta aacatcctac agtgacagag tcagcatccc acagcaaaga ataacttagc
60301 ctgaaaagtc aagagtgcag ccattgagaa atcctgagct agatgcacag actaYatttg
60361 gcaggtgggt tattcttgaa gttctacagct tcaagagctt cctaaaMccc ttttcaaatg
60421 taggctaagt tgagagtctg gttctgggaa aggttccaca gcttttatct gattcttaaa
60481 aagatccgtg gccgaaagag gctgaagaac caggaaaacc aaggggtcag ggaccttggc
60541 tgaaacagat tatttctactc agaaaattat caggaggtga ggctgggtct gtttgcctga
60601 gccagattat agtaaatctt gaaaaccagg cagagatagt tggacatcaa cctgaggtct
60661 gctgtaggac acagaagacc tatttaggaag ataattggcag gacaaaaact gtgtcacaat
60721 atgaatgtct tgaagtggct tctctacttc aactgcccaa gttcccttca atcaggctgc
60781 aaaacttagt attcttatgg agtggactaa actatatcca atcagtcacc aatccatatt
60841 aatttttttt ctcaaaaaata tatttttagtc tttctttccc cgcttatcag ggtcaacacc
60901 ctattctagg cccagctcat cttaactttg gacaactgca aaaccttctc tagtgagtc
60961 ttggcttatt catctttgca aattgagcac tggatatact tcctggcatg tggcagccat
61021 tccgtaaggt acactctcaa gtgagggtgg gaggagaagg ttgtagaatc cccccaagtt
61081 acagtgaaaa caaattacag agtgatttct cattagtccc attatgggga acatgcctca
61141 tccacgttct taatgaggct ccagtaagtc tgtagactgt ttggggctga ggggtgggca
61201 ttggtgtgaa gcagaagtgg tgcttcacag taagtgttgg gataaatcca acaagctgag
61261 ttccccagg actgtgtctt ccacctcccc cgcaagccaca ggaaggtaat cacgtaaat
61321 cagcctgaga aagaggatat gagagttttt tgaagtctac aaacctgtta tttgccgatt
61381 ccaaactcat ttcttaccag actatcctca ggctaaaatt ccttaaaatac caagtttcga
61441 ttatcacttt tcacacttaa actctcaata agtcccaact ctaaatctaa attgtgcagc
61501 taattttttt tggccgagca cgggtgctca cgcctgtaat cccagcactt tgagaggccg
61561 aggcgggcgg agcacagggt cgggagatcg agaccatcct ggctaacatg gtgaaacccc
61621 gtctctacta aaagtacaaa aaaaaaaaaa aaatgccggg cgtggtgcgg acgcctgtag
61681 tcccagctac tccggaggct gaggcaggag aatggcgtga acccgggagg cggagctggc
61741 agtgagccga gatgcacca ctgcactcca gcctgggcga cagaacgaga ctccgtctca

```

61801	aataaataaa	taaataaata	aataaataaa	taaaaataaa	tatatattact	cttcaatttg
61861	gtcaataaat	aaacacgaga	gggtacgaga	aattacttaa	aaaaaaaaa	tccccgcaga
61921	atccaacaat	ctctgaagcc	tttaacacct	aacaaccca	aggggagat	ttcggctctt
61981	ggactgtatt	tcaaagcgat	tggaccccag	accattcccc	cagttactca	aaagatccag
62041	cctctcctca	ggtgtagccc	gctctagaaa	ggccgcctcg	ggagtaccac	ccactcctga
62101	cggggcgggg	agttctctcc	gaccccaccc	gttccaagga	cgcagccctg	tgattggcca
62161	tgacgcgtgc	cagagctgcc	cagtcgcgga	agagcagctg	ttttggcgtc	tgtttgctgc
62221	ggtaggaaca	tggcggatcg	gctcacgcag	cttcaggacg	ctgtgaattc	ggtgaggaat
62281	ttcattcgat	tagcctctgt	ctttctcttt	cttgaggtaa	agcaagggag	gaaaggggcc
62341	caagaggcaa	aacttgRaag	tggggagcgg	actgagagac	agggcttcgg	cataaagcgc
62401	tctctcggga	ggcgccaccg	cgaactcgga	ggtttgaggc	cggctgcggg	gcggtgcttg
62461	aaggtgcggg	agggagcaga	ctctttcgct	tgaggagacc	gggaagatcg	ttcttgtccg
62521	ggtgaggccg	ccaggacttg	tgggtgcgcg	ggtcctgtcg	tgttccctga	ggacagtttt
62581	attttcttgt	cagtgggtgct	tacgggttct	ctttcttttg	ggggttgagg	ggagaagcag
62641	gagctttttt	ctttcctcat	caattttaat	ccaccaaggg	agatctcatg	ctctggatga
62701	ctgggttttt	gtaatgaaat	gtcagccgag	agtattcctt	caccagcgct	ttcaggttgg
62761	tgcgggagcg	atgctgaggt	gaggtgagag	attttctttc	atcttagctc	gttctcttgg
62821	gcccagagta	tccttccatc	tcttgctctt	tggccatccc	tcagaataca	tcagtttggg
62881	gacatcttta	tgatagggaa	tcttctaata	ctcgtatgtc	tttccgctta	tttgagtctt
62941	ttttcttttc	tttttttttt	tttttttact	ttggagacag	tgcccttcgc	ccaggcttga
63001	gtgcagtgcc	gcgatctctg	ctcactgcac	cctccgcctc	cggggttcaa	gcgactctcc
63061	tgtctgagcc	aaagtagctg	ggattacagg	cgccctccac	cacgcccggc	taatgtttgt
63121	atttttagta	gagacggggg	ttcaccatgt	tgtccaggct	ggtctcaaac	tcccggcctc
63181	aagcgatcct	cccgcctcgg	cctcccatag	tgctgggatt	acaggtgtga	gccaccgtgc
63241	ccggcctaca	agtccttttt	aatgttctcc	ccaaattcta	ctctctatta	aaggccaagt
63301	ctagtaagta	gctccccagc	tttgttctct	ctggttattt	ttcctccaca	ctatctctga
63361	gtggaaagta	gaagcaaagc	cagaggatgc	gcagtcattg	gcccctaatt	aaggggtcat
63421	gttctgagaa	atgcatgggt	ttggcgataa	cttttttttt	tgttttaata	agtagaagta
63481	caatgtaaaa	taacgataaa	acagtaatac	ataaaccagt	aacaaagtca	tttagtatca
63541	ttatcaagtg	tcattggactg	cacataattg	tatgtgctat	gctttcatac	actggcagtg
63601	tggtaggttg	atttacagca	gcataccac	aagcactaag	gaacgcattg	gcctacagcg
63661	ttagtactgc	tacgatgtca	ctaggcaata	agaatttttc	agctttatta	taaccatagg
63721	ggactgacat	aatatgtgca	tcccgttgac	caagacgtca	gtgggtattc	ccatgcggca
63781	atgggtgtac	ttattaccat	tacgagttag	attgttttct	gctaaatagt	tttccattga
63841	tagttatggt	tctcattgtc	tgattgtgat	ttaatgtcct	agataagggg	ctagagagaa
63901	aacatgctga	taggaaaacc	taagtacctg	gtttccttgc	cagttttttt	tttccctact
63961	gtaagctaata	aacttggcat	ttttcttaata	acatcagaat	ttgtacactt	gaagagtgtg
64021	aatctacttg	tgataaaaact	gatctaccct	gggtgaataa	cactagtttt	taagctgaga
64081	ggaccagcct	ctaagtgttg	aattcaccaa	gataaattct	accctgcagg	gacttgccctg
64141	ggctgccagc	cagccttctc	cagggaacta	ctcctttgag	gttttacgaa	atcaaattgt
64201	aattgatgctg	ctggatagaa	atcttatttt	agggaaaatt	ttgcatattc	atacatgtct
64261	tgtaactaga	aggataataa	ggacagagtg	aatatggcct	taactccagg	ttaggcataa
64321	tgataaattg	attttggaga	aaagccttta	ggataactag	tatgagtgcc	tactgtgtgt
64381	aagcaccatc	atgtttacaa	agagatactc	ctaaaggagg	ttaaaaaatg	tcaaatactc
64441	aggagacatt	taagtgtatg	aaaaacagta	tgtgggttagg	gagttataat	gaattgagta
64501	ttttcagctt	caggttctga	tgtaaatgtg	aatatggagt	gatgtgaacc	aaaaataggg
64561	gaatcttctt	agcattttatc	ctaatacgaa	ttatttgtat	aacttttaac	atttgttcta
64621	ccctctgaaa	accgtgagtg	cagaagctat	gtcgcattca	ctatgtctct	gtagtgttta
64681	tatgcagtac	tttagttatt	ttccataaat	ttttgttgac	taaatggatc	ttaaaagcct
64741	ttgaaggatg	gaaaagatag	acactgtggt	gaataatttt	ttgtgtgtgt	ttcctctatt
64801	acaacaatgc	tataatagga	gaatgtatcc	aatttaaaaa	tgaaaaatgc	tgcttggtatt
64861	aatagagcta	ggatcaggat	ctaggtttgt	atagcatctg	agcacttatt	gctccattac
64921	aattagcaaa	acctgggcaa	agccatttag	tcataagtga	ctttatgact	agttagataa
64981	atatgaagga	gtagtggaaa	acgagtcagt	agagttgact	tagattattg	aaagcttcaa
65041	aacagcctga	gttggagatt	tatcttgaag	acagtgggtg	agtaagctta	aatgttttga
65101	tcagggttaa	ataaataacc	tggtgttaac	gaaaacaaca	agaagactat	tgtaataatt
65161	ctggtgtgag	ataattgggt	gaactaagca	acagtgatag	aaggaaaaga	gaaagatttg
65221	tagtattagc	agaatatgaa	gatcaaagga	aagtaagaat	ctgaagttag	attggttgac
65281	taggagactt	tcaatacaaa	ttaaatatata	aaaagaggaa	acagttgcct	ttgagttctg
65341	tcattgaaac	aatggcagag	aaacccagtg	gaaatactta	ttagaaatag	gaggttttag
65401	aggtcaagac	taggtgagag	aaattggtag	ggaagatagg	tgattttaat	ttggagatgg
65461	taaggaattg	ctattcccaa	tcttttaaaag	tacttaagga	atacttttag	agtacaagag
65521	tttacaaatt	gagtcctgaa	caaatacatt	tctgcttact	taaggtgctt	ttctgagcag
65581	cctcttcctc	actatttctt	tacacttgta	actactccag	tctcagaaaa	ggaatcatat

65641	actttgtgca	tgagaaataa	atttttatta	aacatggtgc	catagatatt	tacaagtcaa
65701	atttgctcct	ttccatcact	atcccaaaaa	agtctttact	tgtggccttac	aataagtatt
65761	ggccaatatt	ctcttatcat	tttactatca	caagaaaaga	aagattattt	tgtatttcaa
65821	ctgtttcaac	tagattttata	tcaaatttta	ctttacctag	attgatacca	aacttagttt
65881	ctttgcagaa	ttaaaacggt	taatacaaat	taataataac	tggcttattt	tatcaacaat
65941	cactcctttc	tgttttattc	ccttgctcag	agattatcat	tacagaaact	tattgttggt
66001	gccattacca	ataaagtcct	taaaactgat	aagtcctttc	taaccttgat	atgttttctt
66061	tggcctaagc	ttgcagatca	gttttgtaat	gccattggag	tattgcagca	atgtggtcct
66121	cctgcctcct	tcaataatat	tcagacagca	attaacaaag	accagccagc	taacctaca
66181	gaaggtaaac	aggttttctt	agcttctcct	agtttgactc	tcacattttt	tgtaatcctt
66241	taaaattaga	ttcttgaga	aattcagcct	attttatttc	taaaattatt	gttgtttgag
66301	aatattacaa	gttttggttg	ttgtaaaaaa	tttacagtca	attcacctaa	catctcataa
66361	ttctggttca	tgtcagataa	tagattgtta	ctcctttaga	ggagaatttg	tattaattta
66421	gctgagtgag	ggaccttaac	aaggaaaggca	tcccatttat	aaatttcttt	attcaggtga
66481	ccactatatt	gtcataataa	atgaaattac	atacattatg	cttaattccc	aatttgagct
66541	aaatgtgtaa	caggattggt	gtatatttta	atcattagaa	tcttcttttt	tttttgaga
66601	cggagtctcg	ctctgtcacc	taggtcggag	tgcatgatg	caatcttggc	tcactgcaac
66661	tttcgcctcc	cgggttcaag	gaattctcct	gcctcagcct	cctgagtagc	tgggattaca
66721	ggctcatgcc	accacgtctg	gctaattttt	gtatttttag	tagagatggg	gtttcaccat
66781	tgtggtcagg	ctggtctcaa	actcctgacc	tcgtgatccg	cctgccttgg	cctcccaaag
66841	tgctgggatt	acaggcgtga	gccaccagc	ctggccgaat	cattagaatc	ttaggcgaga
66901	aaaagctata	tgaagtttct	ggtttcatta	acttaaaatt	tgtttctgaa	aacttgaggt
66961	tttctaatat	cctaattctag	cagtatcttt	ctctagagta	tgcccagctt	tttgcagcac
67021	tgattgcacg	aacagcaaaa	gacattgatg	ttttgataga	ttccttacct	agtgaagaat
67081	ctacagctgc	tttacaggta	agcctctatt	cctttgagaa	ttttaccagt	aatagagaat
67141	tttgaattaa	aggaggtaga	tttagtacct	tttgcctgtg	tccattttgt	ctgtcataac
67201	aacatacctg	agacagtaat	ttataaagaa	cagaaattta	ttttctcaca	gttctggagt
67261	ctgggaagtc	caagatcaag	gcatttatta	ctcactgtgc	atgagtaatg	aatttttatt
67321	aaacatggtg	ccatgttttag	ggccatgttg	agggccttct	tcctgtttcc	tcacatggct
67381	gaaggcagaa	gggcaaaaaga	gcaagagctc	cctcttcaac	ctcaagccat	attttaaggg
67441	tactaatctc	ctctatgaga	tctctgcctt	catgacttaa	tatcctctta	aaggcccctt
67501	ccctcttaat	accatcacat	tggggattag	attttaatgt	atgaattttg	gaggggatgc
67561	attcaaatca	tagcaccttt	tgttataata	ttcaaatttc	cattataagg	aaacctctta
67621	ttttcctcaa	ttttgtctct	tttctagtg	aagaatgaga	ggagggtggc	attctacaat
67681	tggagtagaa	taatggtgtc	taaacgtaat	gtgtaaggga	gtcacttaaa	gatacttaga
67741	gatctctaatt	tgaaaagtc	ttctttttaa	taattcacat	ggggaatttt	cttaaaagt
67801	gtctgatttt	tagaatttct	ttactgaaaa	acatatcaca	atgattgcaa	agtcattccag
67861	atttaaaaat	aagtacatct	gtgacagctt	ctctttctaa	actctaaaga	ttttaaaatt
67921	tgtgtcttat	aaaggctgct	agcttgata	agctagaaga	agaaaacat	gaagctgcta
67981	catgtctgga	ggatgttggt	tatcgaggag	acatgcttct	ggagaagata	caaagcgcac
68041	ttgtgatatt	gtcacagtca	cagctgaaga	caagaagtgg	tacctatagc	cagctctctc
68101	cagactcata	gcacagtgg	ataccatgtg	gctgagaaaa	gaactgtttg	agtgccatta
68161	agaattctgc	atcagactta	gatacaagcc	ttaccaacaa	ttacagaaac	attaaacact
68221	atgacacatK	accttttttag	ctatttttaa	tagtcttcta	ttttcactct	tgataagctt
68281	ataaaaatca	gattgaatca	gctttaaagc	atcataccat	catttttttaa	ctgagtgaag
68341	ttattaaggc	atgtaataca	ttaatgaaca	taataataagg	aaacatatgt	aaaattctgt
68401	tatgacataa	tttatgtctc	cattttgttg	tattggccag	tactttttaca	aatcaaaaca
68461	tctctcaagc	caaaggagaa	gacagtaaga	acagacataa	gggacatttt	agtttRggct
68521	agtgtcctgc	ctcttagagg	tggcattgtg	taaatctgag	ccttgagcaa	gaaagttttg
68581	gaaattgtgt	tgcttttaag	aaatagagtt	agtgtggctg	gataagaaa	tcacatttat
68641	gcaaatgttt	cttctgctag	aacctcatat	ctgttctagt	tcatctccac	gtttatttta
68701	ccaagagatc	ctctgtcaag	agaatttctt	ggctgttggt	aaagaatttt	tctacatcct
68761	gaactatttt	tcctattttc	tttcaccttg	ctttttttca	tccttaattt	gcttgtcatc
68821	acaatgaagt	atgttttttg	aatcatcaat	tctttctcat	tctccattta	tctgaagggt
68881	cttttgccct	tattaacctc	tcaactttta	ttgctgtatt	ctagtccagc	cagaacatac
68941	ttccactaca	tcagttaact	ggcatcattt	cacctcagtt	tattattgct	aaataaata
69001	ttgtgtcagc	attttcaact	atggttattc	atccaaccct	tggatctctt	tgagYctact
69061	gataatctcc	actggaaagg	tggaaattgaa	atgtggtcca	catttttaact	agctattttc
69121	tggggctgta	tttttcagaa	tatgctgagc	tacatttgct	gcaatctgtt	gctattcaga
69181	gttaaatggt	caggagaaaa	caggaaacat	agaacactct	gcctgttatt	tttggtgtaa
69241	tcaagctttt	ccacagttct	tgaaaagtac	tatgtttcaa	atttcaggaa	caccagcgtt
69301	agctgtaaaa	gttgcagcaa	tttattggct	agtcatagaa	aatttttgaa	cttttaactg
69361	tatttttaatt	gatgtttatt	aaaacactt	tgctatcaga	tatttggcat	aatctgttac
69421	tcttcattat	agttttgggg	ggagagaaga	ttcagtcaga	aaacttatct	aaagtacct

69481	agtattataa	aggagtcaaa	aagggtacaaa	gagaaaaaggt	caagacattt	ttcaaatgag
69541	ggaaaaactaa	caggattttat	cactagtaaa	cctgctctaa	aagaattcaa	gggaagcttt
69601	ttaaaaagaa	gggaagttat	agcagaagga	aacttagaat	ggcaggaata	aagaaggcat
69661	aatgtatagg	gtaaaataaa	tagacttctc	ttgaggtttt	aaaaattaca	tttgttattt
69721	gaaagaaaaa	aattaacggt	gttgatgtg	attctctgta	gaggatatac	agtttttttt
69781	tggtgttctt	gtttctgttt	ttttaagggtg	aagtctctgt	cacccaagct	gaagtgcagt
69841	tctgtgatca	tggctcactg	cagcttcacc	ctgggttcag	gtgatcctcc	cacttcagcc
69901	tcttcagtaa	ctgggactac	aggcatgtac	taccacgtcc	agctaatttt	ttttttcttt
69961	tttttttaga	gatgggggtct	cactctgttg	cccaggctaa	tgccaaactg	ctaacctcaa
70021	gatattaata	ctatacagta	gactataaga	aattaagtat	ttgtatcatc	cctagagcga
70081	cacagatttt	tccttaaaaa	actacagatt	attttcaaaa	gcattaataa	attaagggtg
70141	aatactaaaa	aagattcagg	taaccacagga	acgagaaaca	gaataaaaca	atagtaaacc
70201	taaattttat	tttatttcat	taatacagca	ataattaaat	gtcaatgatc	taagctgatt
70261	taaatctaaa	aactggataa	gaagctgtgg	cttgtaatta	aaaatagacc	tttattcact
70321	agacctacag	tttgcttttt	tcatatgtta	agaaaagttt	aatatcctgc	ttctctattt
70381	cagtttctag	ggcaccacta	tcagttaacc	ctaaagattt	ctgtgggtga	aacctatttg
70441	attaccatct	tgctctgtctg	ccagttatgg	taaccaaact	ctctttgttt	ctggatgcta
70501	aatctgtcca	cttgtgcctt	aactacatcc	taactccgt	catcttcatt	gaatttaagg
70561	agccaatttc	tttttttgag	atggagtttc	actcttgccc	aggctggagt	gcaatggcac
70621	gatcttggct	caccgcaacc	tcagcttctc	gggcttaagt	gattctcctg	cctcagcctc
70681	cgaattagct	gggattacag	gcaggcacca	ccatgcctgg	ctaattttgt	atttttagta
70741	gagacggggt	ttctccatgt	tggctcaggct	ggctctgaac	tcccaacctt	agggtgatcca
70801	ccgcctcag	cctcccaaag	tgctgggatt	acaggcatga	gccactgcac	ccggccagga
70861	gccaatttca	atgggagtat	cctttactat	cattcctgac	ctgtgatatt	ggccattaga
70921	gcctttcaag	aatggtacct	cctgcccagtg	atttttttta	aagccttaat	gaagagagtg
70981	tccctgggat	ccttccagag	aagataggat	gtgagtgaag	aagttatgta	taataatttc
71041	accccaaaat	ttctatgtcc	ttgataccat	tccctgtata	ccaaggaagg	gtattctggc
71101	atttcacttt	gatttaatat	aagtaaacc	tgtgtctggg	cattagttaa	caaaccaagt
71161	gactacaagt	gataagtaac	ttttaccata	gactgtcaat	taccttggca	atggaaacat
71221	tattgtcttt	taatccagtc	agatcacaga	acaaattcca	gatttccatc	cttaagattc
71281	tgtgtctctg	ctaccacttc	tggtgcaaaa	acatgtcagt	gtgaggtaga	gtcaattggaa
71341	tgtagccatg	aaaaaaaaaa	tgctactaca	aacattacag	gaataaggaa	tttgatcgaa
71401	gaattagcat	ttacacaaaa	tatgggagga	gttagggaag	tgaaagaatg	aaaggagaaa
71461	gaggataaag	aaaacactaa	ccacatcacc	ctaaagccct	tattgtctaa	tgagggttatt
71521	gtagtgccag	ttaagtccag	gtcagccaag	gcgtgcgtac	cccttggccc	cacttggaaa
71581	gctgtgtgga	tatgactggt	ggatttaggg	gttctcaaag	ccaaaatgga	tttatcagca
71641	atgaccttac	tgccctgtac	ccagctttga	gaggaccaaa	aacgacggat	gtgtcttgca
71701	ttccctggtc	cctctttctt	cttttacc	tatctaagtt	gtttaaagga	gaaaaatcct
71761	taggttgaag	ccatattctc	tctattcttg	aaggaccaca	gactctacgc	ccctgttgaa
71821	attttctgga	agcctatggg	aaaggaggaa	tgaaaagctt	cgtgatgtct	ttgaatacag
71881	gtgcctaagt	tactatcgta	actgggtccag	taaggagaag	gggtatctga	atttcaactt
71941	gggggctggg	caaggtaacc	agtgaggtgg	ggagataaat	agaacttggg	agctaagacc
72001	cttcagggtg	gcttgatgat	gcattgtgat	catggcttct	acttcagatg	tggtgtgga
72061	atagatatcc	ttatacttgt	aattccccta	tacacaaatg	ccagcaagaa	tagtcccatt
72121	gaggacatgc	taaatgtaga	gaggtagtag	tatgacatgg	aaaacactcc	ccaacttgtc
72181	tccctgtctc	ctgcttatgt	ggtacatggc	cactgactac	agtattaaac	tttaacttgg
72241	ccctgtgctt	ctggaaaaaca	gtgacagtta	agaaatccct	ccttcttggt	acatttcagg
72301	aaatggctta	ttgcaaagaa	ccactagaac	tgtgagtaac	accttcattt	acctatgata
72361	aagccagact	cagaccctcc	aaatcccatt	ctttgcctca	aaataattag	ttgaactatt
72421	tgtacccata	gactaatctg	gacaaaatgc	cctttaactt	gacccaattt	tatgcttctc
72481	ccttccccgt	aagccactgg	actttgatcc	accttcagcc	tgagccagca	ctggactgtg
72541	gaatagtccc	tccttttccc	ctcctgaaaa	ttgactagcc	accaggaagg	acattccctg
72601	ctSagctgtc	ccattaacgt	cacctgtccc	tctcctccac	gttagttttt	ttctagcctg
72661	tttattcctc	cctgtaagag	aaaaagagaa	aagcctat	ctatctgatc	tttgagatgc
72721	ttgcagatct	tacctttgaa	gcattctcct	tattgcaata	gtctccctga	ccctattgca
72781	atagttcctc	tctccctggt	tgtattccac	cattcaggct	gctatatata	gcaaaatacc
72841	ataagctgat	agcttataga	caatataaat	ttacttccca	cagttctgaa	gccccgggaag
72901	tcaaagatca	aagtgtctgg	aggctctggtg	tctggtgggg	gcacctctc	actgtcctca
72961	catggtagaa	gaagcaaggg	aaactctttg	gcctctttag	aggactaat	cccactcatg
73021	aggcctttac	cttcttgacc	tagtcttttt	tcaaaggctc	gacccatcat	cattcgtatg
73081	ttaggatttc	aacgtacaaa	ttgtgtgggt	tgggagggat	acaagcattc	agagcatagc
73141	actattttta	taatcatttc	aaatagagtc	attctatacc	tagtctggat	ttttgtttgt
73201	tttacttgac	ctcatttcact	tgggtaggaa	aaaaatctca	tggactctca	ttgatagccc
73261	caggggtaag	gttggggcaa	aaagagcaga	ggcactcatt	tactgtgagg	taagaaatct

73321	gtttcagatc	accttgtgtg	caattcagga	taaaattaat	aaagatgaat	agtaaaaatc
73381	caatactaag	agataaaaact	caggagtcaa	aagaagaaat	cagaaagatt	agagcataat
73441	tcagaatgta	atgaaaacaa	actatgcata	gaaacttgtg	gtattcggcc	aagacagcat
73501	gtaggagaaa	atgcaagtat	tacctttctt	gcttgtatta	ggataaaaaag	ttgaaaatta
73561	acaagctaag	tatctaactg	aaagagaaaag	aaagaaataa	taaaggtaac	agcagaaatt
73621	aagaaaacaa	ccatactgta	gaaaaaacaa	agcccaaagt	tgagattcaa	aaagagtatg
73681	gctgaaaaaa	gagttttgac	tattgattga	atgtatttaa	aattttaatg	gtccttttaa
73741	tccttcaaac	ttaatccttt	aaacttcaaa	aaaagtttaa	atttggctaa	cgcagtgact
73801	cacacctgta	gtcccagcac	tttgggaggc	aaaggtgggt	ggatcacctg	aggtcagggtg
73861	ttcaagacca	gcctggccaa	catggagaaa	ccccgtctct	actaaaaata	caaaaaatta
73921	gccaggtgtg	gtgggtggcg	cctgtaatcc	aatcccagct	actcgggagg	ctaaggcagg
73981	agaatcggta	cagcctggga	ggcggagggt	gcagtgaagc	gagatcacgc	cattgcactc
74041	cagcctgggc	aacaagggca	aaactccgtc	tcaaaaaaaa	agtttaatac	ataatttaat
74101	aaattttaat	agataaacaa	ggtctcactc	ttgtcttcca	ggctgcagtg	cagtgcacaca
74161	atcacagttc	attgcagttg	tgacctcctg	ggctcaagct	atcctcccac	ctcagcctcc
74221	cgagttagctg	ggatggcagg	cttattccac	cagctaattt	ttgtattttt	agttagatgtg
74281	gggttttgca	gtgttgccca	ggcttatctc	aaactcctgg	gctcaaccga	tctgcccatt
74341	tcgacctccc	aaagtgtctg	gattacatgc	atgagccacc	acaccgatt	tgactcaata
74401	gtcaaaaactc	ttttttggaa	acggaagttc	gcgcttgctg	cccgggctgg	agtgcgggtg
74461	cgtgatctcg	gctcactgca	acctccgtct	cccgggtcca	agtgtctctt	atgcctcagc
74521	ctcctgagta	gctgggatta	caggcaactg	tcaccacgcc	tggtctaattt	ttgtattttt
74581	cgtagagaca	gggtttttacc	atgtttacca	ggctagtctt	gaattcctga	ccttaggtgc
74641	tcgcctgccc	cctggcctcc	ctaagtgtctg	ggattacagg	cgtgagcaac	tgcgcccagc
74701	tgtcaaaaact	cttaattagc	ccggtttgggt	agtgcattgcc	tgaagttcca	gctactcagt
74761	tggtgaggatc	gcttaagcct	gggaggtaca	ggctgcgggtg	agccatgatc	atgacaccac
74821	acaccagcct	ggtgacagac	ttgttgacag	accctgtctc	aaacaaaaata	aaacaacacag
74881	gcagtaataa	aaccttctag	ccaccaagcc	tagagagttt	tattaatgag	aatagtgaat
74941	tgtgagtttt	gctaaatatg	cgttttcttt	gaattccagt	tttatacaga	aagtataaag
75001	ctcttgtcaa	aattttacag	taacagtacc	aaaattaaaa	ataactaaat	ctcttttttt
75061	ttttttggat	taaggtgatt	tgttatttct	ttttttttta	attatacttt	aagtttttagg
75121	gttatgtgca	acaatgtgca	ggttagttac	atatgtatac	atgtgccatg	ctggtgtgct
75181	gcacccgtta	actcgtcatt	tagcattagg	tatatctcct	aatgctatcc	ctccccctc
75241	ccccaccccc	acaacagtcc	ccagagtgtg	atgttcccc	tcctgtgtcc	atgtgttctc
75301	actgttgaat	tcccatctat	gagtgagaac	atgcggtgtt	tggttttttg	tccttgccat
75361	agtttgcgtga	gaatgatgat	ttccaatttc	atccatgtcc	ctacaaaagga	catgaactca
75421	tcattgttta	tggtgcata	gtattccatg	gtgtatatgt	gccacatttt	cttaatccag
75481	tctatcattg	ttggacattt	gggttagttc	caagtctttg	ctattgtgaa	tagtgccgca
75541	atagacatac	gtgtgcatgt	gtctttatag	cagcatgatt	tatagtcctt	tgggtgtata
75601	cccagtagtg	ggatggctgg	gtcaaatggt	attttagatt	ctagatccct	gaggaatcac
75661	cacactgact	tccacaatgg	ttaaactagt	ttacagtccc	accaatagtg	taaaactatt
75721	cctatttctc	aacatcctct	ccagcaactg	ttgtttcctg	acttcttaat	gattgccatt
75781	ctaactggtg	tgagatgata	tcccattgtg	gttttgattt	gcatttctct	gatggccagt
75841	gatgatgagc	attttttcag	tgtcttttgg	ctgcataaat	gtcttctttt	gagaagtgtc
75901	tgttcatatc	ctttgcccac	tttttgatgg	gggtgtttgg	tttttcttgg	taaatttgtt
75961	ggagttcact	gtagattctg	gatattagcc	ctttgtcaga	tgagttaggt	gcgaaaattt
76021	tctcccattc	tgtaggttgc	ctgttcaccc	tgatggtagt	ttcttttggc	gtgcagaagc
76081	tctttagttt	aatttagatcc	catttgtcaa	ttttggcttt	tattgccatt	gcttttggtg
76141	tttcagacat	gaagtccttg	cccattgccta	tgtcctgaat	ggtaatgtct	aggttttctt
76201	ctagggtttt	tatggtttta	ggtctaacgt	ttaagtcttt	aatccatctt	gaattaattt
76261	ttgtataagc	tgtaaaagaag	ggatccagtt	tcagctttct	acatatggct	agccagtttt
76321	cccagcaccg	tttatttaaat	agggaaatcct	ttccccattg	cttgtttttc	tcaggtttgt
76381	caaagatcag	atagttgtag	atatgtggcg	ttatttctga	ggcctctgtt	ctgttccatt
76441	gatttatatc	tctgttttgg	taccagtacc	atgctgtttt	ggttactgta	gccttgtagt
76501	atagtttgaa	gtcaggtagt	gtgatgcctc	cagctttgtt	cttttggctt	aggattgact
76561	tgagagatgtg	ggttcctttt	tggttccata	tgaactttta	agtagttttt	ttccaattct
76621	gtgaagaaaag	gcattggtag	cttgatgggg	atggcattga	atctataaat	taccttgggc
76681	agtatggcca	ttttcacaa	attgattctt	cctaccgatg	agcatggaat	gttcttccat
76741	ttgtttgtat	cctcttttat	ttccttgagc	agtgttttgt	agttcttctt	gaagaggtcc
76801	ttcacgtccc	ttgtaagttg	gattcctaag	tattttattc	tctttgaagc	aattgtgaat
76861	gggagttcac	tcatgatttg	gctctctgtt	tgtctgttat	cgggtgataa	gaatgcttgt
76921	gatttttggta	cattgatttt	gtatcctgag	actttgctga	agttgcttat	cagcttaagg
76981	agattttggg	ctgagacaat	gggttttctt	atacatacag	tcatgtcgtc	tgcaaacagg
77041	gacaatttga	cttctctctt	tcctaattga	atacccttta	tttccttatc	ctgcctaatt
77101	gccctggcca	gaacttccaa	cactatgttg	aataggagtg	gtgagagagg	gcatccctgt

77161	cttgtgcccag	ttttcaaagg	gaatgcttcc	agttttttgcc	cattcagtaa	acatggaaaa
77221	atagcaaatt	agagacagt	ggtatttaaca	attgcatata	tcacaataaa	tgtatgttta
77281	atctcaggaa	tgcatggcag	ttctgtcttt	agaaaatgta	ttttttgtat	ttttatcaat
77341	ttagagatgc	ccttttaaaa	aaacatatta	caaattggaa	tattcaaagg	gcagctgttt
77401	tattttcttc	cttgctgaaa	cataaaataa	tgacgtgttt	taaaatttat	agaatctttg
77461	atttgatcaa	atatggtaat	gtaatttgc	attctaacag	atcaaaatca	gaatcagtat
77521	gatcatctat	cttgatttgc	gagaagcatt	tgataatatt	caaaatccat	ttatgaaaaa
77581	actcttgcca	gaggagaaac	agcaaagaat	ttttaaaata	taaaggggat	gtacaaaact
77641	gtagtaacca	tcttactgaa	agatcaaact	tgttcccttt	cagcttggga	ataagacaga
77701	gcttcctatt	aagaccacca	catctactca	tctttaagct	ggtgattctc	accaacacat
77761	ctaagcatga	aagaaataaa	aattataaga	taaaaaggaa	agaaaactct	cagtcattaa
77821	agatgatata	attacataga	aaatcaagaa	gattccatgg	ataggcttcc	caaaaaacaa
77881	tatgtaacat	accagcaaca	aatggtgaaga	aaaacacgtt	gttttttttt	taaagtacat
77941	tttcagtgc	attaaataat	aaggattcag	gaatacaacg	aatataacaa	aatgtgaaga
78001	tatctaggga	tgcgataaac	aaaaaaatta	taaagcttca	tttaaaacat	ttttaaacaa
78061	aataagtaga	agtaataacc	atctttatgg	aacagagtca	gtattataaa	tggtgcagtt
78121	ttcttcaact	ggtccataga	caacacaatt	ttaactaaga	ttccaattgg	gatttctgtg
78181	gaacttgatg	agttagggcc	aaaagtaatc	aagatattca	tgaggaagga	taacagaatg
78241	gcaaaacttg	ctttaataat	ttgtgcaggg	atagacaaat	actgtacacc	agtctctcaa
78301	tagggagatc	aaaaccaa	tcacacagac	atgaaaagtt	gatttatgac	agatgatgtt
78361	gcagtccttg	gggaaagaat	ggaaatgggtg	ctagaataat	tggtttctat	atgtatttct
78421	atatatttat	atgccattac	attaaagtca	atgcctatgg	agtttgggag	gagtgtacaa
78481	aatgttcttg	aattagatgt	gatgtttgca	caactctgtg	aatgtactaa	gaaccactga
78541	attgtatatt	tttaaagcat	ggtatgtgaa	taatacctca	ataaagctgt	tataaaaaat
78601	attttaaaaa	tcccagcact	ttgggaggct	gaggcgggcg	gatcatgagg	tcaggagttc
78661	aagaccagcc	tgaccaacat	gttgaaaccc	catctctatt	aaaagtacaa	aaattagcca
78721	gacgtggtg	cacgtgcctg	taactaagc	tactcaggtg	gctgaggcag	gaaaatcgct
78781	tgatccctg	aggcggaggc	tgccgtgagc	cgagatgatg	cgcttggcac	tccatccagc
78841	ctgggcaaca	gagcgagatt	ccatctcaaa	aaaaaaaaaa	aagtaaaaaa	tcaatgacgg
78901	ccaggcgag	tggtctcacac	ctgtaatccc	agcactttgg	gaagccgagg	tggcagatca
78961	cctgaggtca	ggagttcgag	accagcttga	ccaacatgga	gaaaccccg	ctctactaaa
79021	agtacaaaat	tagccgggtg	tggtggtgca	tgccgtgaat	cccagctact	ggcgaggctg
79081	aagcaggata	atcacttgaa	cccgggaggc	ggcggttgtg	gtgagccaag	aacgcgccat
79141	tgccctccag	cctgggcaac	aagagcgaaa	ttccatctca	aaaaaaaaaa	aaaaaaaaatc
79201	aatgccagac	agactgaaaa	cttaacatgc	aaagggcaag	tattgaagac	caaacagata
79261	gcaatatttc	accacaagag	ggcagtttga	ccctgaactt	tggtgtgtgt	ttgttttag
79321	atggagtctt	gctctgttgc	ccaggctgga	gtgcagtggc	gtgatcttgg	ctcactgcaa
79381	ccctccacct	cccggttcca	agtgtattacc	ctgcctcagc	ctcccgagta	gctgggacta
79441	caggcacgtg	ccaccatgaa	cgactaattt	tttgtgtttt	tagtagagac	ggggtttcac
79501	cgtgttagcc	aggatggcct	tgatctcctg	accttgtgat	ctgcccacct	cagctttcca
79561	aagtgcctgg	attacaggcg	tgaactttga	ttttatacac	gaagtgtgtg	gaggaaatgg
79621	ttggaaagga	tggttctgtga	ctcaaaagac	tactggggat	taactagctt	gatagcagtg
79681	gccacaagtt	aagattaccc	tggaatccag	tcaaagctgg	tgttatgatt	ctagtttgag
79741	ctagtatctt	cctggagcca	gtaacatgaa	tatgatgacc	aatgtttat	ttcgtctgaa
79801	aacaaaagat	tatggacaca	acttcgtggt	atgtgtccat	ttaggcccac	cgtaaagaag
79861	tgaagcactg	gcatcaacca	caaagaaagt	tatgtgagca	ctttctctgg	atgataaatt
79921	atcttctatg	tgcatgtgca	tgtgtgtgtg	atgttgacag	gtcacacatc	agaaactgcc
79981	acttcatagc	cgaaggatag	tcctatatgg	aaacgcaata	tgaaagtgtg	attaataata
80041	taggcttttg	tgccctcctg	ccccggcatt	gccactgaat	tgtgtgactt	tgggcaagtt
80101	acctatgttt	cagtgaact	gtaaaatgag	gataatatta	gtaattacct	catggtattt
80161	tttttcttct	ttttttta	tatactttaa	gttctagggt	acatgtgcac	aacgtgcagg
80221	tttgtttacat	atgtatacat	atgccatgtt	ggtgttctgc	acccattaac	tcgtcattta
80281	cattaggtat	atctctta	gtatccctc	ccccatcccc	ccacccacg	acaggcccca
80341	gtgtgtgatg	tttcccttcc	tgtgtccaag	tgttctcctt	gttcaattcc	cacctatgag
80401	tgagaacatg	cagtgtttgg	ttttctgtcc	ttgcgataat	ttgctcagaa	tgttggtttc
80461	catgtccctc	catgtcccta	caaaggacat	gaactcatcc	ttttttatgg	ctgcatagta
80521	ttccatgggt	tatatgtgcc	acattttctt	aatccagtct	atcactgatg	gacatttggg
80581	ttggttccaa	gtctttgcta	ttgtgaatag	tgccgcaata	aacatacgtg	tgcatgtgtc
80641	tttatagcag	catgatttat	aatcctttgg	gtatataccc	agtaatggaa	tggctgggtc
80701	aaatggattt	tctagtctta	gatccctgag	gaatcaccac	actgtcttcc	acaattgttg
80761	aactagttta	cagtcccacc	aacaggtgaa	aagtattcct	gtttctccac	ttcctctcca
80821	gcacctgttg	tttcttgact	ttttaatgat	tgccattcta	acgggtgtga	gataaacttc
80881	tccgagctaa	aagaggaagt	tcgaacccat	cagaaagaag	ctaaaaacct	tgaaaaaaga
80941	ttagatgaat	ggctaactgg	aataaccagt	gtagagaagt	ccttaaatga	tctgatggag

81001	ctgaaaatga	tggcacgaga	actacgtgac	aaatgcacaa	gcttcagtag	ccgattttgat
81061	caactggaag	aaaggggtatc	agtgtattgaa	gatcaaacga	atgaaatgaa	gtgagaagag
81121	aagtttagag	aaaaaagagt	aaaaagaaat	gaataaaagcc	tccaagaaat	atgggagtat
81181	gtgaaaagac	caaattctatg	tctgtattagt	gtacctgaaa	gtgatgggag	gaatggaacc
81241	aagttggaag	acactctgca	ggatattatc	caagagaatt	tccccaacct	agcaaggcag
81301	gccaacattc	aaattcagga	aatacacaga	acaccacaaa	aatactcctc	aagaagagca
81361	actccaagac	acataattgt	gagattcacc	aaagttgaaa	tgaaggaaaa	aatgttaagg
81421	gcagccagag	agaaagggtcg	ggttaccac	aaagggaagc	ccgtcaaaact	aacagcagat
81481	ctctcggcag	aaactctaca	agccagaaga	gagtgggggc	caatattcaa	cattctttaa
81541	caaaagaatt	ttcaaccag	aattttacat	ccagccaaac	taagcttcat	aagtgaagga
81601	aaaataaaat	cctttactga	caagcaaatc	tgagagactt	tgctaccacc	aggcctgcc
81661	tacaagagct	cctgaaggaa	gcataaaca	tagaaaggaa	caaccagtac	cagccactgc
81721	aaaaacatgc	caaaatgtaa	agaccatcaa	tgctagaaag	aaactgcatt	aactaacgag
81781	caaagtaacc	agctaaccatc	ataatgacag	gatcaaatc	ccacataaca	gtattaacct
81841	taaatgtaaa	tgggctaaat	gctccaatta	aaagacacag	actggcaaat	tggataaaga
81901	gtcaagacc	atcagtgtgc	tgtattcagg	agaccctct	catgtgcaga	ggcacacata
81961	ggctcaaaat	aaagggagg	agggaatct	accaagcaaa	tggaaaaaca	aaaaaggtag
82021	gggttgcaat	cctatctctg	ataaaacaga	atataaacca	acaaagatga	aaagagacaa
82081	ggccattata	taatggtaaa	tggatcaatt	caacaagaag	agctaactat	cctaaatata
82141	tatgcaccca	atacaggagc	accagattc	ataaagcaag	tccttgagga	cttacaagaa
82201	gacttagact	cccacacaat	aataatggga	gactttaaca	ccccactgtc	aactatagac
82261	agatcaagga	gacagaaggt	taacaaggat	atccaggact	tcacctctgc	accaagggga
82321	cctaatagac	atctacagaa	ctctccacca	caaatacaaca	gaatatacat	tcttctcagc
82381	accacattgc	acttattcca	aaattgagca	catagtggga	agtaaagcac	tcctcagcaa
82441	atgcagaaga	acagaaataa	caaactgtct	ctcagaccac	agtgcaatca	aattagaact
82501	caggattaa	aaactcactc	aaaaccactc	aactacatgg	aaactgaaca	acctgctcct
82561	gaatgactac	tgggtacata	acgaaatgaa	ggcagaaata	aaggtgtttt	ttgaaaccaa
82621	tgagaacaaa	gacacaacgt	accagaatct	ctgggacaca	tttcaagcag	tgtgtagagg
82681	gaaattttata	gcactaaatg	cccacaagag	aaagcaggaa	agatctaaaa	ttgacaccct
82741	aacatcacaa	ttaaaagaac	tagagaagca	agagcaaaaca	cattcaaaa	ctagcagaag
82801	gcaagaaata	ctaagatca	gagcagaact	gaaggaaata	gaaacacaaa	aaaccttca
82861	aaaaatcaat	gaatccagga	gctgggtttt	tgaaaagatc	aacaaaattg	atagaccact
82921	agcaagtcta	ataaagaaga	aaagagagaa	gaatcaaaaca	gatgcaataa	aaaatgataa
82981	aggggatatc	accaccgatc	ccacagaaat	acaaactacc	atcagagaat	actgtaaata
83041	cctctgtgca	aataaaactag	aaaatctaga	agaaatggat	aaattcctgc	acacatacac
83101	cctcccaaga	ctaaatcagg	aagaagtgtga	atccctgaat	agaccaataa	caggctctga
83161	aattgaggca	ataattaata	gcctaccaat	caaaaaaaga	cctcatggta	tttttatgga
83221	atgtaaatca	gtttatagat	gagcaggagt	ctagtatcat	cctttaagac	atatctggtt
83281	tttatagggt	cagaccgggtg	aattaagtag	gcacttgatg	gggatcacag	tgcctgcaac
83341	ttttgggtct	ttaaggatga	ctttaacgtc	taccatagtc	ccaggcaata	ttgtttttaa
83401	tttacaatat	aatTTTTTTT	cattttttaca	attatttata	atatttgtct	tttatgtaca
83461	atattgtcaa	tgccttatct	ttttaaaatt	tactatatgt	gctgagacag	gatgagaggg
83521	acagtttcag	agtctccact	tggctttccc	cacttccgta	gctcttacac	cattggttaa
83581	gaaaccagtg	tggggggtgt	gctaactaca	aagtcagtga	ccagggaat	aatcactgat
83641	gggtcatgaa	accagtgaac	ctactgtgag	tgggatcttt	gcaaggaccc	catgtatcat
83701	ctggcaacca	taacctttt	ccccacaag	gtggccatga	taatgccct	ttgaattaat
83761	tctgcattaa	ttcttaggta	ttaatgtcaa	aacagactct	gccaccaaca	aactcctcaa
83821	aatgccttgg	cttgtctctt	tcccttgtat	acagtaacct	gagtaaatgg	ccacaggtct
83881	ctttgggata	ggtatgagag	aatcattgta	aacacttgca	ttatagggtt	tttctcacgc
83941	aagatcctca	ttgttgttca	gtcattgggt	tctgaatctg	aaaaaggctc	caagttctga
84001	aacagggaag	acatcatgac	tttttggttg	gaagtgtctc	caggatttta	gtcatctatc
84061	tttgttgtgt	atcatatggc	atgtactctt	tgtgtctggc	ttctttcatt	aagtacaatg
84121	atTTTgagat	ttacacagat	tgttagtgca	tcgatctttt	tatttttgaa	tagcatccca
84181	tttaattgat	tccattgatt	gcaatgctgc	ctgtctctct	tgcccttagg	gactcatggt
84241	ctcttaactc	tcatagctct	ccgagggtcg	gaaactccat	tacttgctac	tcagactttg
84301	ctgtttattac	agtaattaa	cctattttgg	ttctgatggg	ttagatttgc	caccaggcac
84361	cttttgtttg	gggatcctgt	tattcccat	gccatcatgt	agcgtagttt	ggtaccaaca
84421	tcaaccatca	gacttagcca	ctgggtgtct	gaaccagctc	tcacaagatg	gatttattct
84481	catcaactct	cacaagattg	ttacattttt	tgaagtttta	tgaaccagtt	gttacccaag
84541	atgTtgaatg	taatatgaca	ggagaattgat	ccatccttat	gtcgatatac	tctcccttat
84601	ccaacattat	tctccttgac	tgcattttca	tgttccagtc	tcattgtgtat	tttctcagct
84661	cttgccagta	catcttaact	aggacctgca	gcgccttttc	ctccataaca	gaccagcac
84721	ccccaagctg	tgTtatatca	tgattgatcc	taatttttta	ttgggtagcc	agaatgtgag
84781	gtgggggttg	cattactctg	ttctcacatt	gctataaaga	aatacctgcy	attaggtaat

84841	tgataaaaga	ggtttaattg	gctcacgttt	ctgcaggctg	tacaggaagc	acagtgctaa
84901	caagtgcctg	gcttctgggg	aggcctcagg	aaacttacaa	ccatggcaga	agggtgaagg
84961	ggagcaggca	tatcacatgg	ccagagcagg	agcaagaaaa	tgagggggag	atgcttttaa
85021	atgaccagat	cttacaagag	ctcactcact	gttgtgatga	cagtaccaag	acgatgggtc
85081	taagccattc	atgagaaaatc	tgccaccatg	atccaatcac	ctctgaccat	ggcctacctc
85141	caacattggg	gattttttat	atgagattttg	ggtggggaca	cacatccaaa	gtatatcagg
85201	tgtgtatcct	gggcagaggt	gggtgggtgg	ggggtggcat	ttatcttgca	aagtagagga
85261	cttggcatca	tcttccagga	aatgggtaga	gctaaccctt	aagagaggga	gaggagaagg
85321	ctgcttgtgc	cagcccggag	gaatcagggg	aagtgcgatt	caagattttc	agcatattaa
85381	ccaagaggac	ccattccaag	tcttaagatc	acagtccctc	ccatcaatac	cctccctgac
85441	cttgatgaag	ccaacttgcc	tgagttgaga	atcataatatt	tctagagctt	gttccactact
85501	cttggaaat	aaggctgaat	ctgaccccca	gttttttctg	cctccagact	acaagagatg
85561	agcatctgct	gccaaagagg	ccttctagct	ttcacacttt	gccttcagtc	gatgattaat
85621	acccttagcc	tttcatattcc	cccccaaagt	attcattgat	agcatttagc	aatgatcata
85681	ttccactttt	tttgtttaca	gatttattga	agtttaattg	atgacaataa	actacatatg
85741	tgtttgtatg	tatgtgtgtg	cagtgcacaa	tttgctaagt	ttttgtttgt	ttttgttttt
85801	gagacaggg	ctcgctttgc	cacacaggct	ggagtgcagt	ggcacagtca	tagctcactg
85861	caacctcagc	cttccacatt	tggttaagttt	tgacatRtat	ctatgaaacc	attatcaaaa
85921	tcaagacaac	aaacattttca	tcaaccccaa	aagtttccctc	ctgctccctt	tagttttatc
85981	cttcccttcta	tccccaggta	actcctgatt	tgctttctga	taggttagtt	tgcattttct
86041	agaattttat	ataaatggaa	taatacagca	tgtattcttt	tttttttttg	gtctggctac
86101	tttcatatag	tataatgatt	ttgagattta	tgtatctcgt	tggtgtatca	atctttttat
86161	tgttgaatag	tgtatcccat	ttaatagatc	tatcaaatgt	gtttatccat	tgatctgttg
86221	atggacattt	gggctatttc	caggtttggtg	ctattatgaa	taaagctgct	atgaatattc
86281	atatacaagt	ctttgtgtgg	acatgttccc	atatttcttg	aataaatatg	tagaagtggg
86341	atggctggga	tatttggcaa	aagtgtttta	agttttcaag	aaagtgccca	tctattttcc
86401	agtcatttga	ctcttttgca	ttctcatcag	cagtgtagga	gagttccagc	tgccccacac
86461	cttcaccagc	agtgggcaaa	gtcttttaac	tttagccatt	ctagtgggtg	tataatggta
86521	actcatttga	gttttttcat	ttcattttta	aattgaggta	caattgacat	gcaatgaact
86581	gcacctattt	aaactgtaca	attttagatg	tttgacatat	gtatatacca	atgaaaccaa
86641	cacaatcaag	atgataaata	tattctacac	cctcaaaaagt	tttctcatat	ccctatgaaa
86701	ttctatcttc	tctcccttcc	ttgctcctga	ttccctaccc	cagataacca	ctaactcttc
86761	ttctgtcatt	agtttgcatt	ttctagagtt	ttatacaagt	ggaatctaca	gtatgtactt
86821	tatatgtatc	aataggctcat	acctttttgt	tacttacatc	tttgttactg	tgacatattt
86881	cactgtatgg	ctaaaccatt	attagttttac	ctggtgacag	acatttgggt	tgtttctaat
86941	ttttgccaaa	cacaaataaa	gtggctatga	acattttacat	acaagtattt	gtatggactt
87001	aagctttaat	ttctcttagg	taaataccta	aggagttaga	tggtctggatc	atataatagg
87061	tgtatatagt	atttaacttt	tttttataac	actgtaaaac	tgtttcatag	tggttgtatg
87121	gaaataaaaa	tgagaaacaa	aataaaacaa	atgtatacca	aaacatgttt	aatgtttcac
87181	cttcaagaaa	agggagcact	ttacaatgag	tgagaaaata	tatatccaca	ttttgaaaaa
87241	aactccaaga	taaaaggcat	tttgacaaca	gatatttgga	aatattttcca	tcgtttacaag
87301	agttttattgt	caaaaatgat	gtatgtccat	ctataaaaact	ctcctatctg	ttgctaaaaac
87361	aaattttatct	gctctatttt	taaatatttt	caaaaagtta	gttttaaac	tatttgttta
87421	aaatatgaaa	atgcaaaacc	ttttatttgt	tccaagagta	atttattgac	atcgaagaaa
87481	atataaattt	attactttaa	caaaaactct	ccttttttagc	tcctcttttt	tgttttgttt
87541	ttataactgg	acagataata	attgtacata	ttcatggggg	acatagcgat	gtttcaatac
87601	atatactgta	gtgattagat	cagggttaatt	aacatatctt	tttaattggg	gaatggagtg
87661	aaaaattaat	actctgattt	agtaagtga	gctaactctt	ccatttgaat	ccatacttct
87721	ctgcaggatc	tttttttagct	ttgagagcca	ttaacaccaa	atattgaagt	aaactaaact
87781	tataaccaga	ccatcatatc	attttatctc	ataaaatgaa	gcaagtattc	ctgtcacttc
87841	tggaacaca	actgaacagt	attgtttact	agtgaaaaaa	aattgagttt	tcaagtgaac
87901	attggaattt	tggaaaactt	gtatcctccc	tatgagcttg	acagcttccc	aatatttaga
87961	gactttttatg	atatattttat	ggtgagacca	gtgataatat	taacaattgt	tatttttgaa
88021	atattgctta	atgaaaccca	tcaataccaa	cccaaatttt	ataaaaaaaa	agcatattta
88081	aaaacatttc	ttggggaggc	atgggtggct	atgcctataa	tcccagcact	ttggggaggc
88141	gaggcaggca	gattacttga	ggccagaagt	tcaagaccac	cctggccaac	atgggtgaac
88201	ctccatctct	acttaaacaa	aaaaacagaa	ttcttgcaat	ataggtgacc	ctcataaatc
88261	acataactct	tctgagagta	aaagaggtgc	attttttggt	agagttatcc	agacttttta
88321	cattaattat	attttatatt	ttcagatccc	aaattggcaa	aagttggaga	atttttctta
88381	aattacaatt	ttcttctgcc	agagccaaaa	atgtaagaaa	actcctagat	accatggtta
88441	tatccaatat	cttaaaatta	ttgttctcag	cacaatgaag	aaattattga	gaacaagtcc
88501	cttccaaaga	aaaaaaaaatc	caataagaga	ttaaatatgt	attgtaattt	tagaagacag
88561	ttaagttgca	tcctcacata	gaatgctctc	tcagacactg	gcttatttcta	aatgcgtcag
88621	caagtaatca	aagagtacag	taaatgccta	ttcttctcta	gaacactgct	ggaacagatt

```

88681  tccccctaac cctgggctaag gctctcttta tttatttagat ggattaaact taattccctt
88741  gatttgtact tatttctaata acagttgata cataggaatt cagaaaggct ttttcagact
88801  tgagatttaa gtgcaatgtc agcttttttt gttttttttt cttttttcag tttggctaaa
88861  gcaaagcaag ggaattgttg ccaagtagag cacactatgc agaacctggg acatatagga
88921  tgactcagca ttgcatTTTA gtctcatgac tttgctggga atcagatgca gctatcgaga
88981  ataccagtag cattggcctt aggttttagta gccatgagtt gagtttagga aaggcagaat
89041  cttactaagc aaactcctga ttattgtctg aagaacttct tggccaaagc attcttgaca
89101  aatcatcctt gaaaaactct ttactatatg gaaaacgtaa agaacaatgg gtgagagcac
89161  aggctctggg tttgaacctt ggctatacca ccacttacac atcatgtgac cgtggttaag
89221  tcccatcctc actctgtgtc tcagttacca tatatacaaa ataacaataa catactgcta
89281  aatacttact aaatccttta catttattat gaatataaac cacatgatcc tatgtggctt
89341  atattaataa taaatataaa ggatttagtc tggtaactaa tgaccagtaa atgtttatta
89401  taattgggtt attttattat ttatttattt agacagagtt tcaactcccat tgcctaggct
89461  ggagtgcagt ggcactcagg tgattctcct gcctcagcct cccaagtagc agggaccaca
89521  ggtacaccac catgcccagc taattttttt tttttttttt tgagatggag tctcactctg
89581  tcaccaggct ggagtgcagt ggcctgatct cggctcactg caacctccac ctctgggtt
89641  caagcaattc tcctgcctca gcctcccaag tagctgggac tacaggcgtg cgccaccatg
89701  ttcagctaata tttgtattt tttagtagaga cggggtttca ccatattgtc caggatgggtc
89761  tcaatctctt gatcttgtga tctgcccacc ttggcctccc aaagtgtctg gataataggc
89821  gtgagccacc acacctggcc tgctaatttt taatattttt tgtagacaca cggtttcacc
89881  atgttgccca ggtgtctctt gaactcctga gctcaagtga tctgcctgac ttggcctccc
89941  aaagtgcgga gattacaggt gtgagtcacc atgccagacc agtttgcttt atttcttaag
90001  cctcacaaca tccctgtagt tattgtgggg taggtatctg ctatggtttg aatgtccact
90061  ccaaaagtca tgttgaaatt taattgccat tgaacagta ctaagagggtg gaacctttaa
90121  gaggtgatta ggccatgaga gctctcataa atgtattaac gacattattg tgggagtggtg
90181  ttcattattg caggacaggg ttcctcataa aaggacgagt ttggcaccct attgtttct
90241  ctctgactct caaggtctct tgcttttcca ccttctgcta tgggatgtca taacaagaag
90301  accttcatca gatgtggta ccttgatctt gaactttccc acctgcagaa gtatgaggac
90361  ataaatttct attcccttaa aattaccagg tctgtagtat tctgttatag cagcacaat
90421  cagactaaga cagtatccct aagggtaaa ttgagttatt tctgagagta aactgaggct
90481  cagaagttaa ttgtccattg acccagagcc tgtgtgtggt agttcctgtg attccaaca
90541  aggcaacttc caatatgtca tctaaatct acaaatttca tttatgtctt gtaagtgtag
90601  aaatcagaaa catcatacaa gctcagtcct ttacgtggtc ttagacttcg ccaaattaag
90661  aagatttata gaattaataa ataagtgtat gaggctgggc gcaatggctc atgcctgtaa
90721  tcccagcact ttgggaggcc gaggtgggtg gatcacctga ggtcaggagt ttgagaccag
90781  cctggccaac atggcaaaac cctgtctcca ctaaaaatac aaaaattagc cagtgtggt
90841  ggtgcatgcc tgtagtccca gctactcagg aggtgaggc aggagaattg cttgaaccgg
90901  ggaagcggag gttgcagtga gccaaagatc caccactgca ctccagtctg ggcaacagaa
90961  cgagactctg tctcaaaaaa aaaaaataaa ttaattaaaa agtgtgtgat attgaacttt
91021  gactcttctt gaacaaagcc tgcatatttt aattgttgag tttatttaat tttctgttga
91081  ctatattttg tttgttgcc ttaatatgat aataggacag agacagtatt gtagagagtt
91141  cagcacagtg gcctgaaaaa ttgaggatac tagaaaagtt cttttatata ttaatctatc
91201  cagcagatat ttattgcagc ccctattgta tgcaggcact gctgtgatac agcaatggat
91261  acatcagaca aggttcttgc tcttatagaa atgacattat agagagagga gaagaacatt
91321  tattccatca ataaatacat aaaaaagaca atttcagata gtgtaagtaa cagctaagaa
91381  aacaataaaa caggatgatt ggatacagga gatgggaaga gatggtgaga agtggggaag
91441  gcttagatat agttaaatgg aaaggcttgt ctttgagggt ggcataatgag ctcagtaactt
91501  cttttttttg agtcagagtc tccctctggt gcccaggctg gagtgagtg gcaYgatctt
91561  ggctctctgc aacctctgcc tcttgggttc aagcgattct catgcctcag ctgggattac
91621  aggcgtgtgc caccatgccc aggtaaattt tgtatttttt tgtagagaca gggttctgcc
91681  gtgttgacca gcctggtctc gaactcctga ccgcaaatga tcagcccacc tcagctctcc
91741  aaagtgtctg gattacaggc atgagccact gcacctggcc ctcaatactt tgatcattag

```

[0241] The following is a human cDNA sequence for TM7SF3.

TM7SF3 cDNA Sequence (SEQ ID NO: 2)

NM_016551 Homo sapiens transmembrane 7 superfamily member 3 (TM7SF3), mRNA

1 ccagccctgg cgtgggcccc gcccgcccca ggcagcaatg gggttcctgc agctgctggt

```

61  cgtagcggtg  ctggcatccg  aacaccgggt  ggctgggtgca  gccgaggtct  tcgggaattc
121  cagcgagggg  cttattgaat  tttctgtggg  gaaatttaga  tacttcgagc  tcaataggcc
181  ctttccagag  gaagctattt  tgcattgatat  ttcaagcaat  gtgacttttc  ttattttcca
241  aatacactca  cagtatcaga  atacaactgt  ttccttttct  ccgactctcc  tttccaattc
301  ctcggaaaca  ggcactgcca  gtggactggt  tttcatcctt  agaccagagc  agagtacatg
361  cacttggtac  ttggggactt  caggcataca  gcctgtccag  aatatggcta  tcctactctc
421  ctactcagaa  agagatcctg  tccctggagg  ctgtaatttg  gagttcgatt  tagatattga
481  tcccaacatt  tacttgaggt  ataatttctt  tgaaacgact  atcaagtttg  cccagcaaaa
541  cctaggctat  gcgagaggcg  tagatccccc  accatgtgac  gctgggacag  accaggactc
601  caggtggagg  ttgcagtatg  atgtctatca  gtattttctg  cctgagaatg  acctcactga
661  ggagatggtg  ctgaagcatc  tgcagaggat  ggtcagtggt  cccaggtga  aggccagtgc
721  tctcaagggt  gttaccctaa  cagctaata  taagacaagt  gtttcttctt  cctccctccc
781  gggacaagg  gtcataata  atgtcattgt  ttgggacccg  tttctaaata  catctgctgc
841  ctacattcct  gctcacacat  acgcttgtag  ctttgaggca  ggagagggt  gttgtgcttc
901  cctaggaaga  gtgtcttcca  aagtgttctt  cactcttttt  gccctgcttg  gtttcttcat
961  ttgtttcttt  ggacacagat  tctggaaaac  agaattattc  tcataggct  ttatcatcat
1021  gggattcttc  ttttatatac  tgattacaag  actgacacct  atcaagtatg  atgtgaatct
1081  gattctgaca  gctgtcactg  gaagcgtcgg  tggaatgttc  ttggtagctg  tgtgggtggc
1141  atttggaatc  ctctcgatct  gcatgctctg  tgttgagcta  gtgctggggt  tcctcatctc
1201  gtcagtact  ttctttactc  cactgggaaa  cctaaagatt  tttcatgatg  atggtgtatt
1261  ctgggtcact  ttctcttgca  tagctatcct  cattccagta  gttttcatgg  gctgcctaa
1321  aatactgaac  atactgactt  gtggagtcat  tggctcctat  tcggtggttt  tagccattga
1381  cagttactgg  tccacaagcc  tttcctacat  cactttgaac  gtactcaaga  gagcgctcaa
1441  caaggatttc  cacagagctt  tcacaaatgt  gccttttcaa  actaatgact  tcattatcct
1501  ggcagtatgg  ggcattgctg  ctgtaagtgg  aattacgta  cagattcgaa  gagagagagg
1561  acgaccgttc  ttccctcccc  acccatacaa  gttatggaag  caagagagag  agcgccgagt
1621  gacaaacatt  ctggacccta  gctaccacat  tcctccattg  agagagaggc  tctatggccg
1681  attaacccag  attaaagggc  tcttcagaaa  ggagcagcca  gctggagaga  gaacgccttt
1741  gcttctgtag  atgcccaggg  gcttggctag  tgtgcctcag  ctttgaggtt  catgcctgga
1801  gtggttcaac  agtctctggt  gcaagtctaa  taagagatca  ggcatatata  tctgttcttt
1861  gcataatatt  atggtgcct  tattgatata  tggttaagggt  gtactagggt  attaggatga
1921  ttgtaagaga  atgagaaaga  tgacaaaag  gttggtggta  gggaggcttt  ttcttatttc
1981  caaatacttg  agaaattacc  ttttggttta  caaatctatg  atcaacttat  tccattaaat
2041  agatacat  aaaaaattaa  aaactgcaaa  aaaaaaaaaa  aaactggtgt  ttctttttat
2101  aaccccttga  aacaagtctc  tcactgagc  ctgtctaaac  tttcgagggt  agtttattat
2161  tgagtcttta  tctgtgacag  tatttgagga  tttagggttt  tgatacttag  gcctttgaat
2221  tttagaatac  aaaaagagaa  gcaagccaga  catggtgggt  cacacctgta  atcccaatac
2281  tgggaagcca  agtggtgagt  atcgcttgag  cccaggaggt  tgagaccgac  atgggcaaca
2341  tgacaagacc  ccattctctac  aaaaaaatt  aaaaaattag  ccaggcatgg  tggcacatgc
2401  ctactcccag  ctcccaggga  gactgagatg  ggaggatccc  tggagccctg  aagattgagg
2461  ctacagtgag  ccttgattgt  gtcactgcac  tccagcttgg  gtgacagaga  cctgtctctg
2521  agaaatt

```

[0242] The following is a human polypeptide sequences for TM7SF3.

TM7SF3 Amino Acid Sequence (SEQ ID NO: 3)

NP_057635 transmembrane 7 superfamily member 3; seven transmembrane protein TM7SF3 [Homo sapiens]

MGFLQLLVAVLASEHRVAGAAEVFGNSSEGLIEFSVGKFRYFELNRPFPPEAILHDISS
 NVTFILFIQIHSQYQNTTVSFSPTLLSNSSETGTASGLVFILRPEQSTCTWYLGTSGIQPVQN
 MAILLSYSERDPVPGGCNLEFDLDIDPNIYLEYNFFETIKFAPANLGYARGVDPPPCDA
 GTDQDSRWRLQYDVYQYFLPENDLTEEMLLKHLQRMVSVQVKASALKVVTLTAND
 KTSVSFSSLPQGQVIYNVIVWDPFLNTSAAYIPAHTYACSFEAGEGSCASLGRVSSKVFF
 TLFALLGFFICFFGHRFWKTELEFFIGFIIMGFFFYILITRLPIKYDVNLILTAVTGSVGGMF
 LVAVWWRFGILSICMLCVGLVLGFLISSVTFFTPLGNLKIHFDDGVFWVTFSCIAILIPVV

FMGCLRILNILTCGVIGSYSVVLADSYWSTSLSYITLNVLKRALNKDFHRAFTNVPFQT
NDFIILAVWGMLAVSGITLQIRRERGRPFPPHPYKLWKQERERRVTNILDPSYHIPPLRE
RLYGRLTQIKGLFQKEQPAGERTPLLL

[0243] Modifications may be made to the foregoing without departing from the basic aspects of the invention. Although the invention has been described in substantial detail with reference to one or more specific embodiments, those of skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the invention, as set forth in the aspects which follow. All publications or patent documents cited in this specification are incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference.

[0244] Citation of the above publications or documents is not intended as an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents. U.S. patents and other publications referenced herein are hereby incorporated by reference.

What is claimed is:

1. A method for identifying a subject at risk of osteoarthritis, which comprises detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the one or more polymorphic variations are detected in a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence in SEQ ID NO: 1-2;
- (b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;
- (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;
- (d) a fragment of a nucleotide sequence of (a), (b), or (c);

whereby the presence of the polymorphic variation is indicative of the subject being at risk of osteoarthritis.

2. The method of claim 1, which further comprises obtaining the nucleic acid sample from the subject.

3. The method of claim 1, wherein the one or more polymorphic variations are detected within a region spanning chromosome positions 27052000 to 27066000 in human genomic DNA.

4. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions selected from the group consisting of 230, 231, 5330, 6334, 11372, 11456, 11501, 13393, 16666, 17596, 19710, 19800, 20297, 20967, 32514, 33159, 37600, 41259, 41329, 50060, 53292, 53393, 56417, 56435, 58847, 59595, 59661, 60355, 60407, 62357, 68230, 68516, 69055, 72603, 73928, 85897 and 91554.

5. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 1 selected from the group consisting of 56435, 59595, 53292, 33159 and 41329.

6. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in linkage disequilibrium with one or more positions in claim 3, 4 or 5.

7. The method of claim 1, wherein detecting the presence or absence of the one or more polymorphic variations comprises:

hybridizing an oligonucleotide to the nucleic acid sample, wherein the oligonucleotide is complementary to a nucleotide sequence in the nucleic acid and hybridizes to a region adjacent to the polymorphic variation;

extending the oligonucleotide in the presence of one or more nucleotides, yielding extension products; and

detecting the presence or absence of a polymorphic variation in the extension products.

8. The method of claim 1, wherein the subject is a human.

9. The method of claim 8, wherein the subject is a human female.

10. The method of claim 8, wherein the subject is a human male.

11. A method for identifying a polymorphic variation associated with osteoarthritis proximal to an incident polymorphic variation associated with osteoarthritis, which comprises:

identifying a polymorphic variation proximal to the incident polymorphic variation associated with osteoarthritis, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-2;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation;

determining the presence or absence of an association of the proximal polymorphic variant with osteoarthritis.

12. The method of claim 11, wherein the incident polymorphic variation is at one or more positions in claim 3, 4 or 5.

13. The method of claim 11, wherein the proximal polymorphic variation is within a region between about 5 kb 5' of the incident polymorphic variation and about 5 kb 3' of the incident polymorphic variation.

14. The method of claim 11, which further comprises determining whether the proximal polymorphic variation is in linkage disequilibrium with the incident polymorphic variation.

15. The method of claim 11, which further comprises identifying a second polymorphic variation proximal to the identified proximal polymorphic variation associated with osteoarthritis and determining if the second proximal polymorphic variation is associated with osteoarthritis.

16. The method of claim 15, wherein the second proximal polymorphic variant is within a region between about 5 kb 5' of the incident polymorphic variation and about 5 kb 3' of the proximal polymorphic variation associated with osteoarthritis.

17. An isolated nucleic acid comprising a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence in SEQ ID NO: 1-2;
- (b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;
- (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;
- (d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and
- (e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);

wherein the nucleotide sequence comprises a polymorphic variation associated with osteoarthritis selected from the group consisting of a thymine thymine repeat at position 56435, a thymine at position 59595, a cytosine at position 53292, a guanine at position 33159 and a thymine at position 41329.

18. An oligonucleotide comprising a nucleotide sequence complementary to a portion of the nucleotide sequence of (a), (b), (c), or (d) in claim 17, wherein the 3' end of the oligonucleotide is adjacent to a polymorphic variation associated with osteoarthritis.

19. A microarray comprising an isolated nucleic acid of claim 17 linked to a solid support.

20. An isolated polypeptide encoded by the isolated nucleic acid sequence of claim 17.

21. A method for identifying a candidate therapeutic for treating osteoarthritis, which comprises:

(a) introducing a test molecule to a system which comprises a nucleic acid comprising a nucleotide sequence selected from the group consisting of:

(i) a nucleotide sequence in SEQ ID NO: 1-2;

(ii) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(iii) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(iv) a fragment of a nucleotide sequence of (a), (b), or (c); or

introducing a test molecule to a system which comprises a protein encoded by a nucleotide sequence of (i), (ii), (iii), or (iv); and

(b) determining the presence or absence of an interaction between the test molecule and the nucleic acid or protein,

whereby the presence of an interaction between the test molecule and the nucleic acid or protein identifies the test molecule as a candidate therapeutic for treating osteoarthritis.

22. The method of claim 21, wherein the system is an animal.

23. The method of claim 21, wherein the system is a cell.

24. The method of claim 21, wherein the nucleotide sequence comprises one or more polymorphic variations associated with osteoarthritis.

25. The method of claim 24, wherein the one or more polymorphic variations associated with osteoarthritis are at one or more positions in claim 3, 4 or 5.

26. A method for treating osteoarthritis in a subject, which comprises contacting one or more cells of a subject in need thereof with a nucleic acid, wherein the nucleic acid comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-2;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(d) a fragment of a nucleotide sequence of (a), (b), or (c); and

(e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);

whereby contacting the one or more cells of the subject with the nucleic acid treats the osteoarthritis in the subject.

27. The method of claim 26, wherein the nucleic acid is RNA or PNA.

28. The method of claim 27, wherein the nucleic acid is duplex RNA.

29. A method for treating osteoarthritis in a subject, which comprises contacting one or more cells of a subject in need thereof with a protein, wherein the protein is encoded by a nucleotide sequence which comprises a polynucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-2;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(d) a fragment of a nucleotide sequence of (a), (b), or (c);

whereby contacting the one or more cells of the subject with the protein treats the osteoarthritis in the subject.

30. A method for treating osteoarthritis in a subject, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the one or more polymorphic variation are detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-2;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

administering an osteoarthritis treatment to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

31. The method of claim 30, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

32. The method of claim 30, wherein the treatment is selected from the group consisting of administering a corticosteroid, a corticosteroid, a nonsteroidal anti-inflammatory drug (NSAID), a cyclooxygenase-2 (COX-2) inhibitor, an antibody, a glucocorticoid, hyaluronic acid, chondrotin sulfate, glucosamine or acetaminophen; prescribing a heat/cold regimen or a joint protection regimen; performing joint surgery; prescribing a weight control regimen; and combinations of the foregoing.

33. A method for detecting or preventing osteoarthritis in a subject, which comprises:
detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:
(a) a nucleotide sequence in SEQ ID NO: 1-2;
(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;
(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;
(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

administering an osteoarthritis prevention or detection procedure to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

34. The method of claim 33, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

35. The method of claim 33, wherein the osteoarthritis prevention is selected from the group consisting of administering a corticosteroid, a nonsteroidal anti-inflammatory drug (NSAID), a cyclooxygenase-2 (COX-2) inhibitor, an antibody, a glucocorticoid, hyaluronic acid, chondrotin sulfate, glucosamine or acetaminophen; prescribing a heat/cold regimen or a joint protection regimen; performing joint surgery; prescribing a weight control regimen; and combinations of the foregoing.

36. A method of targeting information for preventing or treating osteoarthritis to a subject in need thereof, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence in SEQ ID NO: 1-2;
- (b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;
- (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;
- (d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

directing information for preventing or treating osteoarthritis to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

37. The method of claim 36, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

38. A composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and an antibody that specifically binds to a protein, polypeptide or peptide encoded by a nucleotide sequence identical to or 90% or more identical to a nucleotide sequence in SEQ ID NO: 1-2.

39. A composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and a RNA, DNA, PNA or ribozyme molecule comprising a nucleotide sequence identical to or 90% or more identical to a portion of a nucleotide sequence in SEQ ID NO: 1-2.

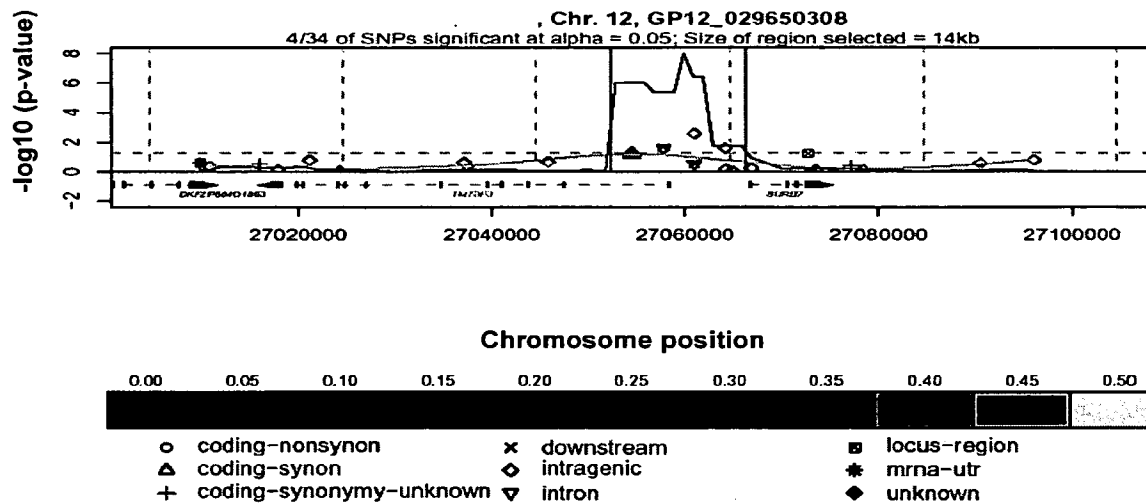
40. The composition of claim 39, wherein the RNA molecule is a short inhibitory RNA molecule.

Abstract of the Disclosure

Provided herein are methods for identifying a risk of osteoarthritis in a subject, reagents and kits for carrying out the methods, methods for identifying candidate therapeutics for treating osteoarthritis, and therapeutic and preventative methods applicable to osteoarthritis. These embodiments are based upon an analysis of polymorphic variations in nucleotide sequences within the human genome.

FIGURE 1

TM7SF3 – DISCOVERY P-VALUES (female only)



Application Data Sheet

Application Information

Application Type::	Provisional
Subject Matter::	Utility
Suggested Group Art Unit::	Not Yet Assigned
CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No
Title::	METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF
Attorney Docket Number::	524593009100
Request for Early Publication?::	No
Request for Non-Publication?::	No
Total Drawing Sheets?::	1
Small Entity?::	Yes
Petition included?::	No
Secrecy Order in Parent Appl.?::	No

Applicant Information

Applicant Authority Type::	Inventor
Primary Citizenship Country::	US
Status::	Full Capacity
Given Name::	Steven
Family Name::	MAH
City of Residence::	San Diego
State or Province of Residence::	CA
Country of Residence::	US
Street of mailing address::	12820 Via Nieve #74
City of mailing address::	San Diego
State or Province of mailing address::	CA

Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Germany
Status:: Full Capacity
Given Name:: Andreas
Family Name:: BRAUN
City of Residence:: San Diego
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 3935 Lago Di Grata Circle
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Germany
Status:: Full Capacity
Given Name:: Stefan
Middle Name:: M.
Family Name:: KAMMERER
City of Residence:: San Diego
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 3825 Elijah Court, Unit 334
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor
Primary Citizenship Country:: US
Status:: Full Capacity

Given Name:: Matthew
Middle Name:: Roberts
Family Name:: NELSON
City of Residence:: San Marcos
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 1250 Calle Prospero
City of mailing address:: San Marcos
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92069

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Sweden
Status:: Full Capacity
Given Name:: Rikard
Middle Name:: Henry
Family Name:: RENELAND
City of Residence:: San Diego
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 7555 Charmant Drive, #1114
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92122

Applicant Authority Type:: Inventor
Primary Citizenship Country:: United Kingdom
Status:: Full Capacity
Given Name:: Maria
Middle Name:: L.
Family Name:: LANGDOWN
City of Residence:: San Diego

State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 3701 Yosemite Street
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92109

Correspondence Information

Correspondence Customer Number:: 25225

Representative Information

Representative Customer Number:: 25225